# LECTURE NOTES

# on

# PROBABILITY and STATISTICS

Eusebius Doedel

# TABLE OF CONTENTS

# SAMPLE SPACES

**DEFINITION** :

The *sample space* is the set of all possible outcomes of an experiment.

**EXAMPLE** : When we *flip a coin* then sample space is

$$\mathcal{S} = \{ H , T \} ,$$

where

$H$ denotes that the coin lands "Heads up"

and

$T$ denotes that the coin lands "Tails up".

For a "*fair coin* " we expect H and T to have the same "*chance* " of occurring, *i.e.*, if we flip the coin many times then about 50 % of the outcomes will be $H$.

We say that the *probability* of H to occur is 0.5 (or 50 %) .

The probability of $T$ to occur is then also 0.5.

**EXAMPLE** :

When we *roll a fair die* then the sample space is

$$\mathcal{S} \;=\; \{\, 1 \;,\; 2 \;,\; 3 \;,\; 4 \;,\; 5 \;,\; 6 \,\} \;.$$

The probability the die lands with $k$ up is $\frac{1}{6}$ , $(k = 1, 2, \cdots, 6)$.

When we roll it 1200 times we expect a 5 up about 200 times.

The probability the die lands with an *even number* up is

$$\frac{1}{6} \;+\; \frac{1}{6} \;+\; \frac{1}{6} \;\;=\;\; \frac{1}{2} \;.$$

**EXAMPLE** :

When we toss a coin 3 times and record the results in the *sequence* that they occur, then the sample space is

$$\mathcal{S} = \{ HHH , HHT , HTH , HTT , THH , THT , TTH , TTT \} .$$

Elements of $\mathcal{S}$ are "*vectors* ", "*sequences* ", or "*ordered outcomes* ".

We may expect each of the 8 outcomes to be equally likely.

Thus the probability of the sequence $HTT$ is $\frac{1}{8}$ .

The probability of a sequence to contain precisely two Heads is

$$\frac{1}{8} + \frac{1}{8} + \frac{1}{8} \quad = \quad \frac{3}{8} .$$

**EXAMPLE** : When we toss a coin 3 times and record the results without paying attention to the order in which they occur, *e.g.*, if we only record the number of Heads, then the sample space is

$$\mathcal{S} \;=\; \Big\{\, \{H,H,H\}\,,\; \{H,H,T\}\,,\; \{H,T,T\}\,,\; \{T,T,T\} \,\Big\}\,.$$

The outcomes in $\mathcal{S}$ are now *sets* ; *i.e.*, order is not important.

Recall that the ordered outcomes are

$$\{\; HHH\;,\; HHT\;,\; HTH\;,\; HTT\;,\; THH\;,\; THT\;,\; TTH\;,\; TTT\;\}\,.$$

Note that

| | | | |
|---|---|---|---|
| $\{H,H,H\}$ | corresponds to | *one* | of the ordered outcomes, |
| $\{H,H,T\}$ | ,, | *three* | ,, |
| $\{H,T,T\}$ | ,, | *three* | ,, |
| $\{T,T,T\}$ | ,, | *one* | ,, |

Thus $\{H,H,H\}$ and $\{T,T,T\}$ each occur with probability $\frac{1}{8}$,

while $\{H,H,T\}$ and $\{H,T,T\}$ each occur with probability $\frac{3}{8}$.

## Events

In Probability Theory subsets of the sample space are called *events*.

**EXAMPLE** : The set of basic outcomes of rolling a die *once* is

$$\mathcal{S} = \{ 1 , 2 , 3 , 4 , 5 , 6 \} ,$$

so the subset $E = \{ 2 , 4 , 6 \}$ is an example of an event.

If a die is rolled *once* and it lands with a 2 *or* a 4 *or* a 6 up then we say that the event $E$ has *occurred*.

We have already seen that the probability that $E$ occurs is

$$P(E) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} .$$

## The Algebra of Events

Since events are *sets*, namely, subsets of the sample space $\mathcal{S}$, we can do the usual *set operations* :

If $E$ and $F$ are events then we can form

| | | |
|---|---|---|
| $E^c$ | the | *complement* of $E$ |
| $E \cup F$ | the | *union* of $E$ and $F$ |
| $EF$ | the | *intersection* of $E$ and $F$ |

We write $E \subset F$ if $E$ is a *subset* of $F$.

**REMARK** : In Probability Theory we use

$$E^c \quad \text{instead of} \quad \bar{E} \ ,$$

$$EF \quad \text{instead of} \quad E \cap F \ ,$$

$$E \subset F \quad \text{instead of} \quad E \subseteq F \ .$$

If the sample space $\mathcal{S}$ is *finite* then we typically allow any subset of $\mathcal{S}$ to be an event.

**EXAMPLE** : If we randomly draw *one character* from a box containing the characters $a$, $b$, and $c$, then the sample space is

$$\mathcal{S} \;=\; \{a\ ,\ b\ ,\ c\}\ ,$$

and there are 8 possible events, namely, those in the set of events

$$\mathcal{E} \;=\; \Big\{\ \{\ \}\ ,\ \{a\}\ ,\ \{b\}\ ,\ \{c\}\ ,\ \{a,b\}\ ,\ \{a,c\}\ ,\ \{b,c\}\ ,\ \{a,b,c\}\ \Big\}\ .$$

If the outcomes $a$, $b$, and $c$, are equally likely to occur, then

$$P(\{\ \}) = 0\ \ ,\ \ P(\{a\}) = \frac{1}{3}\ \ ,\ \ P(\{b\}) = \frac{1}{3}\ \ ,\ \ P(\{c\}) = \frac{1}{3}\ ,$$

$$P(\{a,b\}) = \frac{2}{3}\ ,\ P(\{a,c\}) = \frac{2}{3}\ ,\ P(\{b,c\}) = \frac{2}{3}\ ,\ P(\{a,b,c\}) = 1\ .$$

For example, $P(\{a,b\})$ is the probability the character is an $a$ *or* a $b$.

We always assume that the set $\mathcal{E}$ of allowable events *includes the complements, unions, and intersections* of its events.

**EXAMPLE** : If the sample space is

$$\mathcal{S} \;=\; \{a\ ,\ b\ ,\ c\ ,\ d\}\ ,$$

and we start with the events

$$\mathcal{E}_0 \;=\; \Big\{\ \{a\}\ ,\ \{c,d\}\ \Big\}\ ,$$

then this set of events needs to be extended to (at least)

$$\mathcal{E} \;=\; \Big\{\ \{\,\}\ ,\ \{a\}\ ,\ \{c,d\}\ ,\ \{b,c,d\}\ ,\ \{a,b\}\ ,\ \{a,c,d\}\ ,\ \{b\}\ ,\ \{a,b,c,d\}\ \Big\}\ .$$

**EXERCISE** : Verify $\mathcal{E}$ includes complements, unions, intersections.

## Axioms of Probability

A *probability function* $P$ assigns a real number (the *probability* of $E$) to every event $E$ in a sample space $\mathcal{S}$.

$P(\cdot)$ must satisfy the following basic properties :

- $0 \leq P(E) \leq 1$ ,
- $P(\mathcal{S}) = 1$ ,
- For any *disjoint events* $E_i$ , $i = 1, 2, \cdots, n$, we have

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = P(E_1) + P(E_2) + \cdots P(E_n) \ .$$

## Further Properties

**PROPERTY 1** :

$$P(E \cup E^c) \;=\; P(E) \;+\; P(E^c) \;=\; 1\,. \qquad (\text{ Why ? })$$

Thus

$$P(E^c) \;=\; 1 \;-\; P(E)\,.$$

**EXAMPLE** :

What is the probability of at least one "H" in *four tosses* of a coin?

**SOLUTION** : The sample space $\mathcal{S}$ will have 16 outcomes. (Which?)

$$P(\text{at least one H}) \;=\; 1 \;-\; P(\text{no H}) \;=\; 1 \;-\; \frac{1}{16} \;=\; \frac{15}{16}\,.$$

**PROPERTY 2** :

$$P(E \cup F) \;=\; P(E) \;+\; P(F) \;-\; P(EF)\;.$$

**PROOF** (using the third axiom) :

$$\begin{aligned} P(E \cup F) \;&=\; P(EF) \;+\; P(EF^c) \;+\; P(E^c F) \\ &=\; [P(EF) \;+\; P(EF^c)] \;+\; [P(EF) \;+\; P(E^c F)] \;-\; P(EF) \\ &= \mathrm{P(E)} \;+\; \mathrm{P(F)} \;-\; \mathrm{P(EF)}\;. \qquad (\text{ Why ? }) \end{aligned}$$

**NOTE** :

- Draw a Venn diagram with $E$ and $F$ to see this !

- The formula is similar to the one for the number of elements :

$$n(E \cup F) \;=\; n(E) \;+\; n(F) \;-\; n(EF)\;.$$

So far our sample spaces $\mathcal{S}$ have been *finite*.

$\mathcal{S}$ can also be *countably infinite*, *e.g.*, the set $\mathbb{Z}$ of all integers.

$\mathcal{S}$ can also be *uncountable*, *e.g.*, the set $\mathbb{R}$ of all real numbers.

**EXAMPLE** : Record the low temperature in Montreal on January 8 in each of a large number of years.

We can take $\mathcal{S}$ to be the set of *all real numbers*, *i.e.*, $\mathcal{S} = \mathbb{R}$.

(Are there are other choices of $\mathcal{S}$ ?)

What probability would you expect for the following *events* to have?

(a) $P(\{\pi\})$ (b) $P(\{x \ : \ -\pi < x < \pi\})$

(How does this differ from finite sample spaces?)

We will encounter such infinite sample spaces many times $\cdots$

## Counting Outcomes

We have seen examples where the outcomes in a *finite* sample space $\mathcal{S}$ are *equally likely* , *i.e.*, they have *the same probability* .

Such sample spaces occur quite often.

Computing probabilities then requires counting *all* outcomes and counting *certain types* of outcomes .

The counting has to be done carefully!

We will discuss a number of representative examples in detail.

Concepts that arise include *permutations* and *combinations*.

## Permutations

- Here we count of the number of "*words*" that can be formed from a collection of items (*e.g.*, letters).

- (Also called *sequences*, *vectors*, *ordered sets*.)

- The *order* of the items in the word is important;

  *e.g.*, the word $acb$ is different from the word $bac$.

- The *word length* is the number of characters in the word.

**NOTE** :

For *sets* the order is not important. For example, the set $\{a,c,b\}$ is the same as the set $\{b,a,c\}$.

**EXAMPLE** : Suppose that four-letter words of *lower case* alphabetic characters are generated randomly with equally likely outcomes. (Assume that *letters may appear repeatedly*.)

(a) How many four-letter words are there in the sample space $\mathcal{S}$ ?

**SOLUTION** : $26^4 = 456,976$ .

(b) How many four-letter words are there are there in $\mathcal{S}$ that start with the letter "*s*" ?

**SOLUTION** : $26^3$ .

(c) What is the *probability* of generating a four-letter word that starts with an "*s*" ?

**SOLUTION** :

$$\frac{26^3}{26^4} = \frac{1}{26} \cong 0.038 .$$

Could this have been computed more easily?

**EXAMPLE** : How many re-orderings (*permutations*) are there of the string *abc* ? (Here *letters may appear only once*.)

**SOLUTION** : Six, namely, *abc* , *acb* , *bac* , *bca* , *cab* , *cba* .

If these permutations are generated randomly with equal probability then what is the probability the word starts with the letter "*a*" ?

**SOLUTION** :

$$\frac{2}{6} = \frac{1}{3} .$$

**EXAMPLE** : In general, if the word length is $n$ and *all characters are distinct* then there are $n!$ permutations of the word. ( Why ? )

If these permutations are generated randomly with equal probability then what is the probability the word starts with a particular letter ?

**SOLUTION** :

$$\frac{(n-1)!}{n!} = \frac{1}{n} . \qquad ( \text{Why ?} )$$

**EXAMPLE** : How many

*words* of length $k$

can be formed from

a set of $n$ (distinct) characters ,

(where $k \leq n$ ) ,

when letters can be used *at most once* ?

**SOLUTION** :

$$\begin{aligned} & n\ (n-1)\ (n-2)\ \cdots\ (n-(k-1)) \\ = \quad & n\ (n-1)\ (n-2)\ \cdots\ (n-k+1) \\ = \quad & \frac{n!}{(n-k)!} \qquad (\text{ Why ? }) \end{aligned}$$

**EXAMPLE** : *Three-letter words* are generated randomly from the *five* characters $a$ , $b$ , $c$ , $d$ , $e$ , where letters can be *used at most once.*

(a) How many three-letter words are there in the sample space $\mathcal{S}$ ?

**SOLUTION** : $5 \cdot 4 \cdot 3 = 60$ .

(b) How many words containing $a$ , $b$ are there in $\mathcal{S}$ ?

**SOLUTION** : First place the characters

$$a \ , \ b$$

*i.e.*, *select the two indices* of the locations to place them.

This can be done in

$$3 \times 2 = 6 \text{ ways .} \qquad (\text{ Why ? })$$

There remains one position to be filled with a $c$ , $d$ or an $e$ .

Therefore the number of words is $3 \times 6 = 18$ .

(c) Suppose the 60 solutions in the sample space are *equally likely* .

What is the *probability* of generating a three-letter word that contains the letters *a* and *b* ?

**SOLUTION** :

$$\frac{18}{60} = 0.3 \ .$$

**EXERCISE** :

Suppose the sample space $\mathcal{S}$ consists of all *five-letter* words having *distinct alphabetic characters* .

- How many words are there in $\mathcal{S}$ ?

- How many "special" words are in $\mathcal{S}$ for which *only* the second and the fourth character are vowels, *i.e.*, one of $\{a, e, i, o, u, y\}$ ?

- Assuming the outcomes in $\mathcal{S}$ to be equally likely, what is the probability of drawing such a special word?

## Combinations

Let $S$ be a set containing $n$ (distinct) elements.

Then

a *combination* of $k$ elements from $S$,

is

*any selection* of $k$ elements from $S$,

where *order is not important*.

(Thus the selection is a *set*.)

**NOTE**: By definition a *set* always has *distinct elements*.

**EXAMPLE** :

There are three *combinations* of 2 elements chosen from the set

$$S = \{a\ ,\ b\ ,\ c\}\ ,$$

namely, the *subsets*

$$\{a, b\} \quad , \quad \{a, c\} \quad , \quad \{b, c\} \quad ,$$

whereas there are six *words* of 2 elements from $S$ ,

namely,

$$ab\ ,\ ba \quad , \quad ac\ ,\ ca \quad , \quad bc\ ,\ cb\ .$$

In general, given

$$\text{a set } S \text{ of } n \text{ elements ,}$$

the number of possible subsets of $k$ elements from $S$ equals

$$\binom{n}{k} \equiv \frac{n!}{k!\ (n-k)!} .$$

**REMARK** : The notation $\binom{n}{k}$ is referred to as

"$n$ *choose* $k$ ".

**NOTE** : $$\binom{n}{n} = \frac{n!}{n!\ (n-n)!} = \frac{n!}{n!\ \ 0!} = 1 ,$$

since $0! \equiv 1$ (by "convenient definition" !) .

**PROOF** :

First recall that there are

$$n\ (n-1)\ (n-2)\ \cdots\ (n-k+1) \quad = \quad \frac{n!}{(n-k)!}$$

possible *sequences* of $k$ distinct elements from $S$ .

However, every sequence of length $k$ has $k!$ permutations of itself, and each of these defines the same subset of $S$.

Thus the total number of subsets is

$$\frac{n!}{k!\ (n-k)!} \quad \equiv \quad \binom{n}{k} \ .$$

**EXAMPLE** :

In the previous example, with 2 elements chosen from the set

$$\{a\ ,\ b\ ,\ c\}\ ,$$

we have $n = 3$ and $k = 2$ , so that there are

$$\frac{3!}{(3-2)!} \quad = \quad 6 \quad \textit{words}\ ,$$

namely

$$ab\ ,\ ba \quad ,\quad ac\ ,\ ca \quad ,\quad bc\ ,\ cb \quad ,$$

while there are

$$\binom{3}{2} \quad \equiv \quad \frac{3!}{2!\ (3-2)!} \quad = \quad \frac{6}{2} \quad = \quad 3 \quad \textit{subsets}\ ,$$

namely

$$\{a, b\} \quad , \quad \{a, c\} \quad , \quad \{b, c\}\ .$$

**EXAMPLE** : If we choose 3 elements from $\{a\ ,\ b\ ,\ c\ ,\ d\}$ , then

$$n\ =\ 4\ \text{ and }\ k\ =\ 3\ ,$$

so there are

$$\frac{4!}{(4-3)!}\ =\ 24\quad \text{words, namely :}$$

$$\begin{array}{cccccccc}
abc & , & abd & , & acd & , & bcd & , \\
acb & , & adb & , & adc & , & bdc & , \\
bac & , & bad & , & cad & , & cbd & , \\
bca & , & bda & , & cda & , & cdb & , \\
cab & , & dab & , & dac & , & dbc & , \\
cba & , & dba & , & dca & , & dcb & ,
\end{array}$$

while there are

$$\binom{4}{3}\quad \equiv\quad \frac{4!}{3!\ (4-3)!}\ =\ \frac{24}{6}\ =\ 4\quad \text{subsets}\ ,$$

namely,

$$\{a,b,c\}\ ,\ \{a,b,d\}\ ,\ \{a,c,d\}\ ,\ \{b,c,d\}\ .$$

**EXAMPLE** :

(a) How many ways are there to choose a committee of 4 persons from a group of 10 persons, if order is not important?

**SOLUTION** :

$$\binom{10}{4} = \frac{10!}{4!\,(10-4)!} = 210\ .$$

(b) If each of these 210 outcomes is equally likely then what is the probability that a particular person is on the committee?

**SOLUTION** :

$$\binom{9}{3} \Big/ \binom{10}{4} = \frac{84}{210} = \frac{4}{10}\ . \qquad (\text{ Why ? })$$

Is this result surprising?

(c) What is the probability that a particular person is *not* on the committee?

**SOLUTION** :

$$\binom{9}{4} / \binom{10}{4} = \frac{126}{210} = \frac{6}{10} . \qquad (\text{ Why ? })$$

Is this result surprising?

(d) How many ways are there to choose a committee of 4 persons from a group of 10 persons, if one is to be the chairperson?

**SOLUTION** :

$$\binom{10}{1} \binom{9}{3} = 10 \binom{9}{3} = 10 \frac{9!}{3!\,(9-3)!} = 840 .$$

**QUESTION** : Why is this four times the number in (a) ?

**EXAMPLE** : *Two balls* are selected at random from a bag with *four white* balls and *three black* balls, where order is not important.

What would be an appropriate sample space $\mathcal{S}$ ?

**SOLUTION** : Denote the set of balls by

$$B \;=\; \{w_1\ ,\ w_2\ ,\ w_3\ ,\ w_4\ ,\ b_1\ ,\ b_2\ ,\ b_3\}\ ,$$

where same color balls are made "distinct" by numbering them.

Then a good choice of the sample space is

$$\mathcal{S} \;=\; \text{the set of } \textit{all subsets} \text{ of } \textit{two balls} \text{ from } B\ ,$$

because the wording "*selected at random*" suggests that each such subset has the same chance to be selected.

The number of outcomes in $\mathcal{S}$ (which are sets of two balls) is then

$$\binom{7}{2} \;=\; 21\ .$$

**EXAMPLE** : ( continued $\cdots$ )

(*Two balls* are selected at random from a bag with *four white* balls and *three black* balls.)

- What is the probability that both balls are white?

**SOLUTION** :

$$\binom{4}{2} / \binom{7}{2} = \frac{6}{21} = \frac{2}{7}.$$

- What is the probability that both balls are black?

**SOLUTION** :

$$\binom{3}{2} / \binom{7}{2} = \frac{3}{21} = \frac{1}{7}.$$

- What is the probability that one is white and one is black?

**SOLUTION** :

$$\binom{4}{1}\binom{3}{1} / \binom{7}{2} = \frac{4 \cdot 3}{21} = \frac{4}{7}.$$

(Could this have been computed differently?)

**EXAMPLE** : ( continued $\cdots$ )

In detail, the sample space $\mathcal{S}$ is

$$\Big\{ \begin{array}{lll|lll} \{w_1,w_2\}, & \{w_1,w_3\}, & \{w_1,w_4\}, & \{w_1,b_1\}, & \{w_1,b_2\}, & \{w_1,b_3\}, \\ & \{w_2,w_3\}, & \{w_2,w_4\}, & \{w_2,b_1\}, & \{w_2,b_2\}, & \{w_2,b_3\}, \\ & & \{w_3,w_4\}, & \{w_3,b_1\}, & \{w_3,b_2\}, & \{w_3,b_3\}, \\ & & & \{w_4,b_1\}, & \{w_4,b_2\}, & \{w_4,b_3\}, \\ \hline & & & & & \\ & & & & \{b_1,b_2\}, & \{b_1,b_3\}, \\ & & & & & \{b_2,b_3\} \end{array} \Big\}$$

- $\mathcal{S}$ has 21 outcomes, *each of which is a set* .
- We assumed each outcome of $\mathcal{S}$ has probability $\frac{1}{21}$.
- The *event* "both balls are white" contains 6 outcomes.
- The *event* "both balls are black" contains 3 outcomes.
- The *event* "one is white and one is black" contains 12 outcomes.
- What would be different had we worked with *sequences* ?

**EXERCISE** :

*Three balls* are selected at random from a bag containing

2 *red* , 3 *green* , 4 *blue* balls .

What would be an appropriate sample space $\mathcal{S}$ ?

What is the the number of outcomes in $\mathcal{S}$ ?

What is the probability that all three balls are *red* ?

What is the probability that all three balls are *green* ?

What is the probability that all three balls are *blue* ?

What is the probability of *one red*, *one green*, and *one blue* ball ?

**EXAMPLE** : A bag contains 4 *black* balls and 4 *white* balls.

Suppose one draws *two balls at the time*, until the bag is empty.

What is the probability that each drawn pair is *of the same color*?

**SOLUTION** : An *example of an outcome* in the sample space $\mathcal{S}$ is

$$\Big\{ \{w_1, w_3\} \ , \ \{w_2, b_3\} \ , \ \{w_4, b_1\} \ , \ \{b_2, b_4\} \Big\} \ .$$

The number of such *doubly unordered* outcomes in $\mathcal{S}$ is

$$\frac{1}{4!}\binom{8}{2}\binom{6}{2}\binom{4}{2}\binom{2}{2} = \frac{1}{4!}\frac{8!}{2!\ 6!}\frac{6!}{2!\ 4!}\frac{4!}{2!\ 2!}\frac{2!}{2!\ 0!} = \frac{1}{4!}\frac{8!}{(2!)^4} = 105 \ \text{(Why?)}$$

The number of such outcomes with *pairwise the same color* is

$$\frac{1}{2!}\binom{4}{2}\binom{2}{2} \ \cdot \ \frac{1}{2!}\binom{4}{2}\binom{2}{2} \ = \ 3 \cdot 3 \ = \ 9 \ . \qquad \text{( Why ? )}$$

Thus the probability each pair is *of the same color* is $9/105 \ = \ 3/35$ .

**EXAMPLE** : ( continued $\cdots$ )

The 9 outcomes of *pairwise the same color* constitute the *event*

$$\left\{ \begin{array}{l} \left\{ \{w_1, w_2\}\ ,\ \{w_3, w_4\}\ ,\ \{b_1, b_2\}\ ,\ \{b_3, b_4\} \right\}, \\ \left\{ \{w_1, w_3\}\ ,\ \{w_2, w_4\}\ ,\ \{b_1, b_2\}\ ,\ \{b_3, b_4\} \right\}, \\ \left\{ \{w_1, w_4\}\ ,\ \{w_2, w_3\}\ ,\ \{b_1, b_2\}\ ,\ \{b_3, b_4\} \right\}, \\ \\ \left\{ \{w_1, w_2\}\ ,\ \{w_3, w_4\}\ ,\ \{b_1, b_3\}\ ,\ \{b_2, b_4\} \right\}, \\ \left\{ \{w_1, w_3\}\ ,\ \{w_2, w_4\}\ ,\ \{b_1, b_3\}\ ,\ \{b_2, b_4\} \right\}, \\ \left\{ \{w_1, w_4\}\ ,\ \{w_2, w_3\}\ ,\ \{b_1, b_3\}\ ,\ \{b_2, b_4\} \right\}, \\ \\ \left\{ \{w_1, w_2\}\ ,\ \{w_3, w_4\}\ ,\ \{b_1, b_4\}\ ,\ \{b_2, b_3\} \right\}, \\ \left\{ \{w_1, w_3\}\ ,\ \{w_2, w_4\}\ ,\ \{b_1, b_4\}\ ,\ \{b_2, b_3\} \right\}, \\ \left\{ \{w_1, w_4\}\ ,\ \{w_2, w_3\}\ ,\ \{b_1, b_4\}\ ,\ \{b_2, b_3\} \right\} \end{array} \right\}\ .$$

**EXERCISE** :

- How many ways are there to choose a committee of 4 persons from a group of 6 persons, if order is not important?
- Write down the list of all these possible committees of 4 persons.
- If each of these outcomes is equally likely then what is the probability that two particular persons are on the committee?

**EXERCISE** :

Two balls are selected at random from a bag with three white balls and two black balls.

- Show all elements of a suitable sample space.
- What is the probability that both balls are white?

**EXERCISE** :

We are interested in *the day of the year* each of you is born.

- What is a good sample space $\mathcal{S}$ for this purpose?
- How many outcomes are there in $\mathcal{S}$ ?
- What is the probability of *no common birthdays* in this class?
- What is the probability of *common birthdays* in this class?

**EXAMPLE** :

How many *nonnegative* integer solutions are there to

$$x_1 + x_2 + x_3 = 17 \, ?$$

**SOLUTION** :

Consider seventeen 1's separated by bars to indicate the possible values of $x_1$, $x_2$, and $x_3$, *e.g.*,

$$111|11111111|11111 \, .$$

The total number of positions in the "display" is $17 + 2 = 19$ .

The total number of *nonnegative* solutions is now seen to be

$$\binom{19}{2} = \frac{19!}{(19-2)!\ 2!} = \frac{19 \times 18}{2} = 171 \, .$$

**EXAMPLE** :

How many *nonnegative* integer solutions are there to the *inequality*

$$x_1 \ + \ x_2 \ + \ x_3 \ \leq \ 17 \ ?$$

**SOLUTION** :

Introduce an *auxiliary variable* (or "*slack variable*")

$$x_4 \ \equiv \ 17 \ - \ (x_1 \ + \ x_2 \ + \ x_3) \ .$$

Then

$$x_1 \ + \ x_2 \ + \ x_3 \ + \ x_4 \ = \ 17 \ .$$

Use seventeen 1's separated by 3 bars to indicate the possible values of $x_1$ , $x_2$ , $x_3$ , and $x_4$ , *e.g.*,

$$111|11111111|1111|11 \ .$$

$$111|11111111|1111|11 \ .$$

The total number of positions is

$$17 \ + \ 3 \ = \ 20 \ .$$

The total number of *nonnegative* solutions is therefore

$$\binom{20}{3} \quad = \quad \frac{20!}{(20-3)!\ 3!} \quad = \quad \frac{20 \times 19 \times 18}{3 \times 2} \quad = \quad 1140 \ .$$

**EXAMPLE** :

How many *positive* integer solutions are there to the equation

$$x_1 + x_2 + x_3 = 17 \ ?$$

**SOLUTION** : Let

$$x_1 = \tilde{x}_1 + 1 \quad , \quad x_2 = \tilde{x}_2 + 1 \quad , \quad x_3 = \tilde{x}_3 + 1 \ .$$

Then the problem becomes :

How many *nonnegative* integer solutions are there to the equation

$$\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 = 14 \ ?$$

$$111|111111111|11$$

The solution is

$$\binom{16}{2} = \frac{16!}{(16-2)!\ 2!} = \frac{16 \times 15}{2} = 120 \ .$$

**EXAMPLE** :

What is the probability the *sum* is 9 in *three rolls of a die* ?

**SOLUTION** : The number of such *sequences* of three rolls with sum 9 is the number of integer solutions of

$$x_1 + x_2 + x_3 = 9 \ ,$$

with

$$1 \leq x_1 \leq 6 \quad , \quad 1 \leq x_2 \leq 6 \quad , \quad 1 \leq x_3 \leq 6 \ .$$

Let

$$x_1 = \tilde{x}_1 + 1 \quad , \quad x_2 = \tilde{x}_2 + 1 \quad , \quad x_3 = \tilde{x}_3 + 1 \ .$$

Then the problem becomes :

How many *nonnegative* integer solutions are there to the equation

$$\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 = 6 \ ,$$

with

$$0 \leq \tilde{x}_1 \ , \ \tilde{x}_2 \ , \ \tilde{x}_3 \leq 5 \ .$$

**EXAMPLE** : ( continued $\cdots$ )

Now the equation

$$\tilde{x}_1 \; + \; \tilde{x}_2 \; + \; \tilde{x}_3 \; = \; 6 \;\; , \qquad ( \; 0 \;\; \leq \;\; \tilde{x}_1 \; , \; \tilde{x}_2 \; , \; \tilde{x}_3 \;\; \leq \;\; 5 \; ) \; ,$$

has

$$1|111|11$$

$$\binom{8}{2} \;\; = \;\; 28 \;\; \text{solutions} \; ,$$

from which we must *subtract* the 3 *impossible* solutions

$$(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3) \;\; = \;\; (6,0,0) \;\; , \;\; (0,6,0) \;\; , \;\; (0,0,6) \; .$$

$$111111|| \;\; , \;\; |111111| \;\; , \;\; ||111111$$

Thus the probability that the *sum* of 3 rolls equals 9 is

$$\frac{28 \; - \; 3}{6^3} \;\; = \;\; \frac{25}{216} \;\; \cong \;\; 0.116 \; .$$

**EXAMPLE** : ( continued $\cdots$ )

The 25 outcomes of the event "*the sum of the rolls is* 9" are

$$\{\ 126\ ,\ \ 135\ ,\ \ 144\ ,\ \ 153\ ,\ \ 162\ ,$$

$$216\ ,\ \ 225\ ,\ \ 234\ ,\ \ 243\ ,\ \ 252\ ,\ \ 261\ ,$$

$$315\ \ \ 324\ ,\ \ 333\ ,\ \ 342\ ,\ \ 351\ ,$$

$$414\ ,\ \ 423\ ,\ \ 432\ ,\ \ 441\ ,$$

$$513\ ,\ \ 522\ ,\ \ 531\ ,$$

$$612\ ,\ \ 621\ \ \ \}\ .$$

The "lexicographic" ordering of the *outcomes* (which are *sequences*) in this *event* is used for systematic counting.

**EXERCISE** :

- How many integer solutions are there to the inequality

$$x_1 + x_2 + x_3 \leq 17 \ ,$$

if we require that

$$x_1 \geq 1 \quad , \quad x_2 \geq 2 \quad , \quad x_3 \geq 3 \ ?$$

**EXERCISE** :

What is the probability that the *sum* is *less than or equal to* 9 in *three rolls of a die* ?

# CONDITIONAL PROBABILITY

Giving more information can change the probability of an event.

**EXAMPLE** :

If a coin is tossed two times then what is the probability of two Heads?

**ANSWER** : $\frac{1}{4}$.

**EXAMPLE** :

If a coin is tossed two times then what is the probability of two Heads, *given that the first toss gave Heads* ?

**ANSWER** : $\frac{1}{2}$.

**NOTE** :

Several examples will be about *playing cards* .

A standard *deck* of *playing cards* consists of 52 cards :

- Four *suits* :

  *Hearts* , *Diamonds* (*red*) , and *Spades* , *Clubs* (*black*) .

- Each suit has 13 cards, whose *denomination* is

  $2$ , $3$ , $\cdots$ , $10$ , *Jack* , *Queen* , *King* , *Ace* .

- The *Jack* , *Queen* , and *King* are called *face cards* .

**EXERCISE** :

Suppose we draw a card from a shuffled set of 52 playing cards.

- What is the probability of drawing a Queen ?
- What is the probability of drawing a Queen, given that the card drawn is of *suit* Hearts ?
- What is the probability of drawing a Queen, given that the card drawn is a *Face card* ?

What do the answers tell us?

(We'll soon learn the events "Queen" and "Hearts" are *independent* .)

The two preceding questions are examples of *conditional probability* .

Conditional probability is an *important* and *useful* concept.

If $E$ and $F$ are events, *i.e.*, subsets of a sample space $\mathcal{S}$ , then

$$P(E|F) \quad \textit{is the conditional probability of } E \textit{ , given } F \ ,$$

defined as

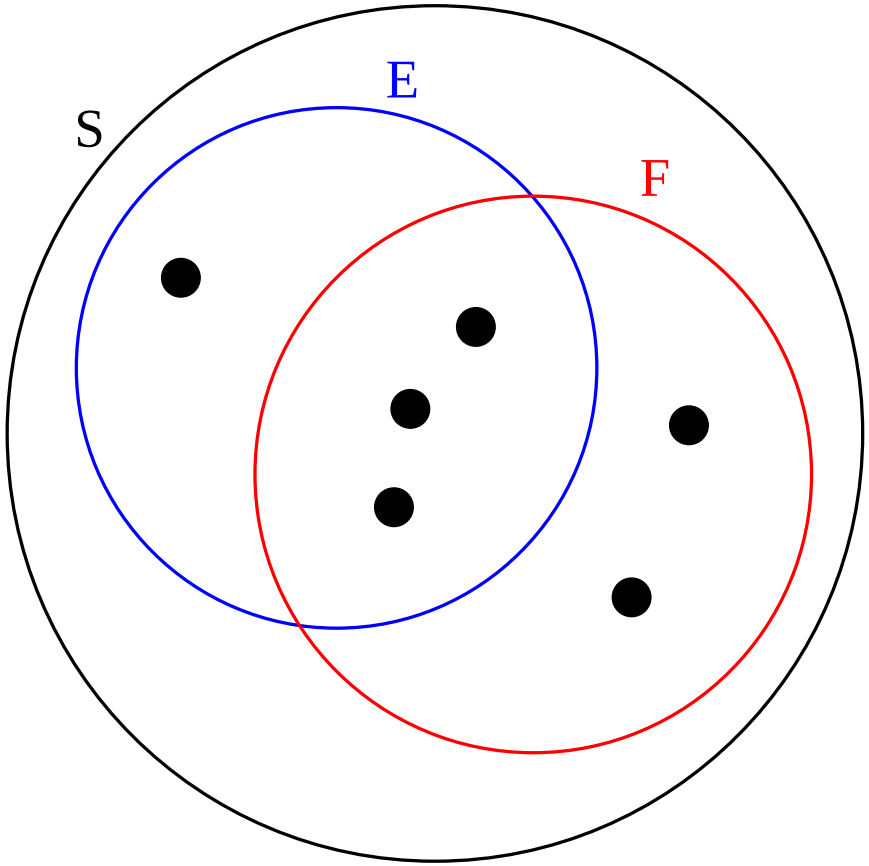
$$P(E|F) \equiv \frac{P(EF)}{P(F)} \ .$$

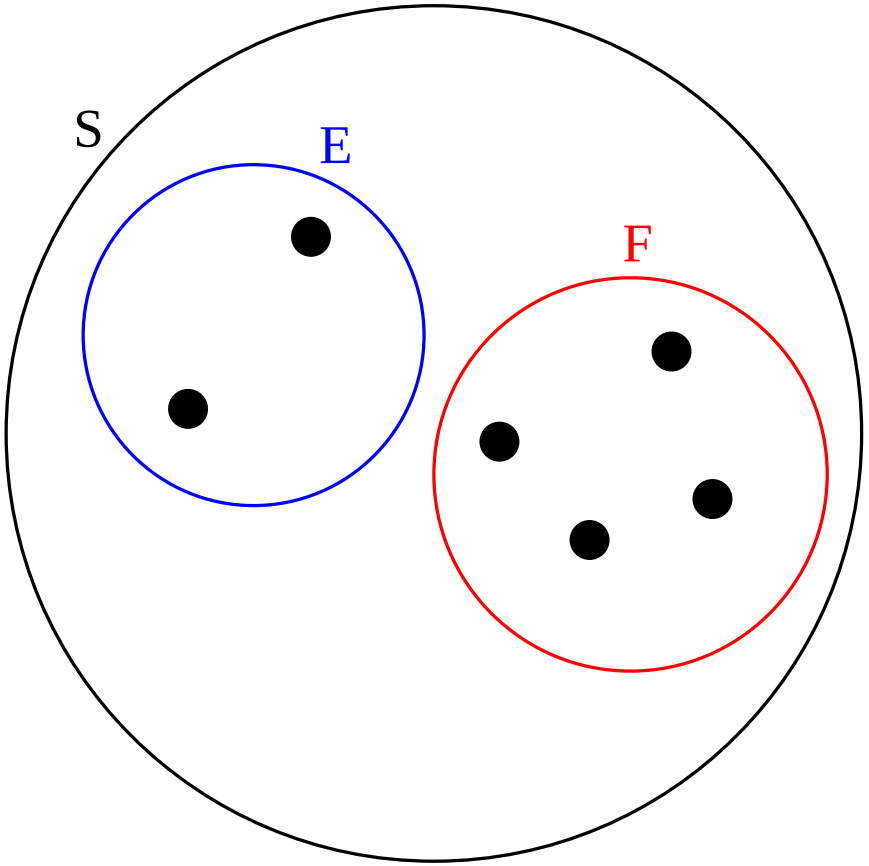
or, equivalently

$$P(EF) = P(E|F)\ P(F) \ ,$$

(assuming that $P(F)$ is not zero).

$$P(E|F) \equiv \frac{P(EF)}{P(F)}$$

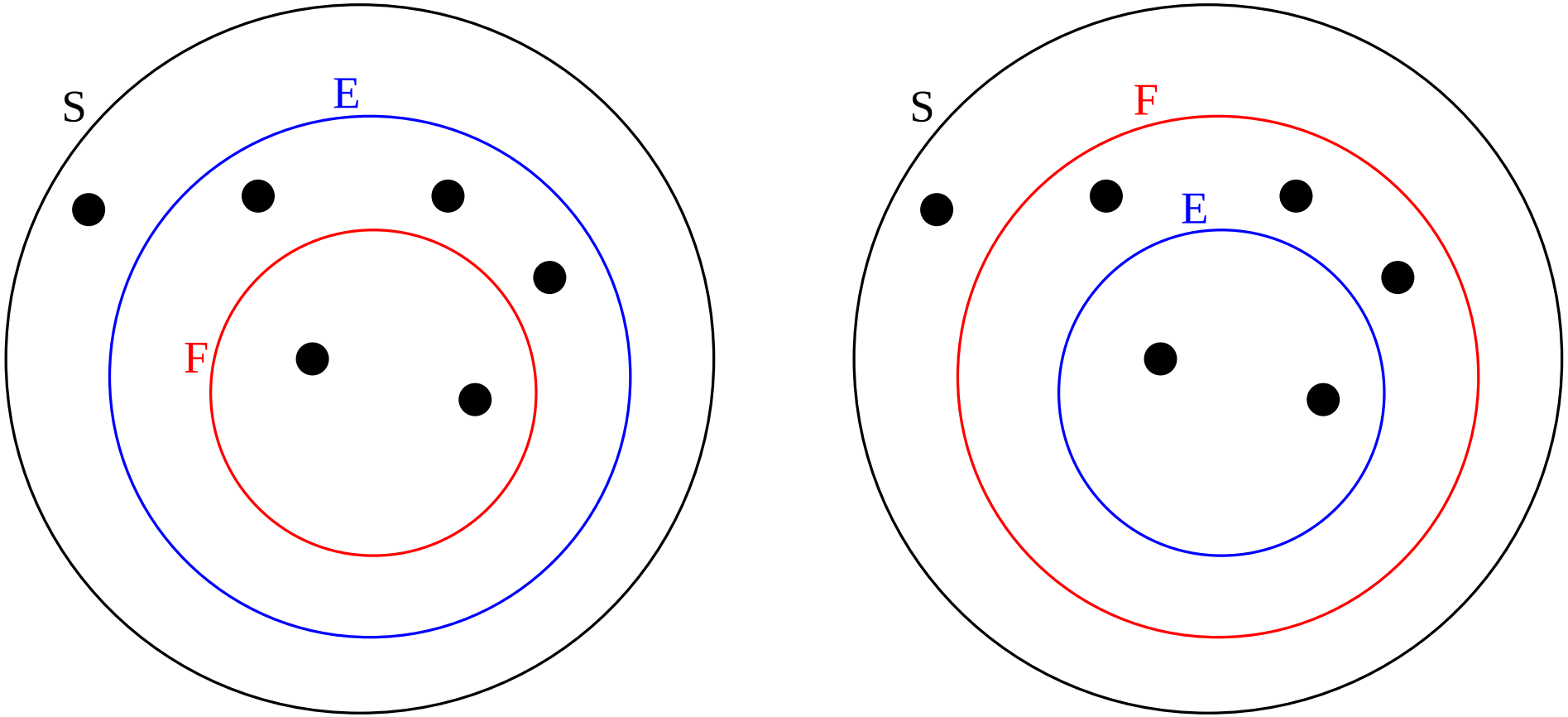



Suppose that the 6 outcomes in $\mathcal{S}$ are equally likely.

What is $P(E|F)$ in each of these two cases ?

$$P(E|F) \equiv \frac{P(EF)}{P(F)}$$




Suppose that the 6 outcomes in $\mathcal{S}$ are equally likely.

What is $P(E|F)$ in each of these two cases ?

**EXAMPLE** : Suppose a coin is tossed two times.

The sample space is

$$\mathcal{S} \;=\; \{HH\ ,\ HT\ ,\ TH\ ,\ TT\}\ .$$

Let $E$ be the event "*two Heads*" , *i.e.*,

$$E \;=\; \{HH\}\ .$$

Let $F$ be the event "*the first toss gives Heads*" , *i.e.*,

$$F \;=\; \{HH\ ,\ HT\}\ .$$

Then

$$EF \;=\; \{HH\} \;=\; E \qquad (\text{ since } \ E \subset F\ )\ .$$

We have

$$P(E|F) \;=\; \frac{P(EF)}{P(F)} \;=\; \frac{P(E)}{P(F)} \;=\; \frac{\frac{1}{4}}{\frac{2}{4}} \;=\; \frac{1}{2}\ .$$

**EXAMPLE** :

Suppose we draw a card from a shuffled set of $52$ playing cards.

- What is the probability of drawing a Queen, given that the card drawn is of *suit* Hearts ?

  **ANSWER** :

$$P(Q|H) \;=\; \frac{P(QH)}{P(H)} \;=\; \frac{\frac{1}{52}}{\frac{13}{52}} \;=\; \frac{1}{13}\,.$$

- What is the probability of drawing a Queen, given that the card drawn is a *Face card* ?

  **ANSWER** :

$$P(Q|F) \;=\; \frac{P(QF)}{P(F)} \;=\; \frac{P(Q)}{P(F)} \;=\; \frac{\frac{4}{52}}{\frac{12}{52}} \;=\; \frac{1}{3}\,.$$

  (Here $Q \subset F$ , so that $QF = Q$ .)

The probability of an event $E$ is sometimes computed more easily

*if we condition $E$ on another event $F$ ,*

namely, from

$$P(E) \quad = \quad P(\ E(F \cup F^c)\ ) \qquad (\text{ Why ? })$$

$$= \quad P(\ EF \cup EF^c\ ) \quad = \quad P(EF) \ + \ P(EF^c) \qquad (\text{ Why ? })$$

and

$$P(EF) \ = \ P(E|F)\ P(F) \quad , \quad P(EF^c) \ = \ P(E|F^c)\ P(F^c)\ ,$$

we obtain this *basic formula*

$$P(E) \quad = \quad P(E|F) \ \cdot \ P(F) \quad + \quad P(E|F^c) \ \cdot \ P(F^c)\ .$$

**EXAMPLE** :

An insurance company has these data :

The probability of an insurance claim in a period of one year is

4 percent for persons under age 30

2 percent for persons over age 30

and it is known that

30 percent of the targeted population is under age 30.

What is the probability of an insurance claim in a period of one year for a randomly chosen person from the targeted population?

**SOLUTION** :

Let the sample space $\mathcal{S}$ be all persons under consideration.

Let $C$ be the event (subset of $\mathcal{S}$) of persons filing a claim.

Let $U$ be the event (subset of $\mathcal{S}$) of persons under age 30.

Then $U^c$ is the event (subset of $\mathcal{S}$) of persons over age 30.

Thus

$$\begin{aligned} P(C) &= P(C|U)\ P(U) + P(C|U^c)\ P(U^c) \\ &= \frac{4}{100}\ \frac{3}{10} + \frac{2}{100}\ \frac{7}{10} \\ &= \frac{26}{1000} = 2.6\% \ . \end{aligned}$$

**EXAMPLE** :

*Two* balls are drawn from a bag with 2 *white* and 3 *black* balls.

There are 20 outcomes (*sequences*) in $\mathcal{S}$ . ( Why ? )

What is the probability that *the second ball is white* ?

**SOLUTION** :

Let $F$ be the event that *the first ball is white*.

Let $S$ be the event that *the second second ball is white*.

Then

$$P(S) = P(S|F)\,P(F) + P(S|F^c)\,P(F^c) = \frac{1}{4}\cdot\frac{2}{5} + \frac{2}{4}\cdot\frac{3}{5} = \frac{2}{5}.$$

**QUESTION** : Is it surprising that $P(S) = P(F)$ ?

**EXAMPLE** : ( continued $\cdots$ )

Is it surprising that $P(S) \;=\; P(F)$ ?

**ANSWER** : Not really, if one considers the sample space $\mathcal{S}$ :

$$\left\{ \begin{array}{llll} \mathbf{w}_1\mathbf{w}_2\ , & \mathbf{w}_1 b_1\ , & \mathbf{w}_1 b_2\ , & \mathbf{w}_1 b_3\ , \\ \mathbf{w}_2\mathbf{w}_1\ , & \mathbf{w}_2 b_1\ , & \mathbf{w}_2 b_2\ , & \mathbf{w}_2 b_3\ , \\ b_1\mathbf{w}_1\ , & b_1\mathbf{w}_2\ , & b_1 b_2\ , & b_1 b_3\ , \\ b_2\mathbf{w}_1\ , & b_2\mathbf{w}_2\ , & b_2 b_1\ , & b_2 b_3\ , \\ b_3\mathbf{w}_1\ , & b_3\mathbf{w}_2\ , & b_3 b_1\ , & b_3 b_2 \end{array} \right\}\ ,$$

where outcomes (*sequences*) are assumed equally likely.

**EXAMPLE** :

Suppose we draw *two cards* from a shuffled set of 52 playing cards.

What is the probability that the second card is a Queen ?

**ANSWER** :

$$P(2^{\text{nd}} \text{ card } Q) =$$

$$P(2^{\text{nd}} \text{ card } Q | 1^{\text{st}} \text{ card } Q) \cdot P(1^{\text{st}} \text{ card } Q)$$

$$+ P(2^{\text{nd}} \text{ card } Q | 1^{\text{st}} \text{ card not } Q) \cdot P(1^{\text{st}} \text{ card not } Q)$$

$$= \frac{3}{51} \cdot \frac{4}{52} + \frac{4}{51} \cdot \frac{48}{52} = \frac{204}{51 \cdot 52} = \frac{4}{52} = \frac{1}{13} .$$

**QUESTION** : Is it surprising that $P(2^{\text{nd}} \text{ card } Q) = P(1^{\text{st}} \text{ card } Q)$ ?

A useful formula that "*inverts conditioning*" is derived as follows :

Since we have both

$$P(EF) \quad = \quad P(E|F)\ P(F)\ ,$$

and

$$P(EF) \quad = \quad P(F|E)\ P(E)\ .$$

If $P(E) \neq 0$ then it follows that

$$P(F|E) \quad = \quad \frac{P(EF)}{P(E)} \quad = \quad \frac{P(E|F) \cdot P(F)}{P(E)}\ ,$$

and, using the earlier useful formula, we get

$$P(F|E) \quad = \quad \frac{P(E|F) \cdot P(F)}{P(E|F) \cdot P(F) \ + \ P(E|F^c) \cdot P(F^c)}\ ,$$

which is known as *Bayes' formula* .

**EXAMPLE** : Suppose 1 in 1000 persons has a certain disease.

A test detects the disease in 99 % of diseased persons.

The test also "detects" the disease in 5 % of healthly persons.

With what probability does a positive test diagnose the disease?

**SOLUTION** : Let

$$D \sim \text{"diseased"} \quad , \quad H \sim \text{"healthy"} \quad , \quad + \sim \text{"positive"}.$$

We are given that

$$P(D) = 0.001 \, , \qquad P(+|D) = 0.99 \, , \qquad P(+|H) = 0.05 \, .$$

By Bayes' formula

$$P(D|+) = \frac{P(+|D) \cdot P(D)}{P(+|D) \cdot P(D) + P(+|H) \cdot P(H)}$$

$$= \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 \; + \; 0.05 \cdot 0.999} \cong 0.0194 \quad (!)$$

**EXERCISE** :

Suppose 1 in 100 products has a certain defect.

A test detects the defect in 95 % of defective products.

The test also "detects" the defect in 10 % of non-defective products.

- With what probability does a positive test diagnose a defect?

**EXERCISE** :

Suppose 1 in 2000 persons has a certain disease.

A test detects the disease in 90 % of diseased persons.

The test also "detects" the disease in 5 % of healthly persons.

- With what probability does a positive test diagnose the disease?

More generally, if the sample space $\mathcal{S}$ is *the union of disjoint events*

$$\mathcal{S} \;=\; F_1 \cup F_2 \cup \cdots \cup F_n \;\;,$$

then for any event $E$

$$P(F_i|E) \;=\; \frac{P(E|F_i)\cdot P(F_i)}{P(E|F_1)\cdot P(F_1) \;+\; P(E|F_2)\cdot P(F_2) \;+\cdots+\; P(E|F_n)\cdot P(F_n)}$$

**EXERCISE** :

Machines $M_1$, $M_2$, $M_3$ produce these *proportions* of a article

*Production* : $\quad M_1 \;:\; 10\,\% \;\;,\quad M_2 \;:\; 30\,\% \;\;,\quad M_3 \;:\; 60\,\%\;.$

The probability the machines produce *defective* articles is

*Defects* : $\quad M_1 \;:\; 4\,\% \;\;,\quad M_2 \;:\; 3\,\% \;\;,\quad M_3 \;:\; 2\,\%\;.$

What is the probability a random article was made by machine $M_1$, given that it is defective?

## Independent Events

Two events $E$ and $F$ are *independent* if

$$P(EF) \quad = \quad P(E)\ P(F)\ .$$

In this case

$$P(E|F) \quad = \quad \frac{P(EF)}{P(F)} \quad = \quad \frac{P(E)\ P(F)}{P(F)} \quad = \quad P(E)\ ,$$

(assuming $P(F)$ is not zero).

Thus

*knowing $F$ occurred doesn't change the probability of $E$ .*

**EXAMPLE** : Draw *one* card from a deck of 52 playing cards.

*Counting outcomes* we find

$$\begin{aligned} P(\text{Face Card}) &= \tfrac{12}{52} = \tfrac{3}{13}\ , \\ P(\text{Hearts}) &= \tfrac{13}{52} = \tfrac{1}{4}\ , \\ P(\text{Face Card and Hearts}) &= \tfrac{3}{52}\ , \\ P(\text{Face Card}|\text{Hearts}) &= \tfrac{3}{13}\ . \end{aligned}$$

We see that

$$P(\text{Face Card and Hearts}) = P(\text{Face Card}) \cdot P(\text{Hearts}) \quad (= \frac{3}{52}\ )\ .$$

Thus the events "*Face Card*" and "*Hearts*" are *independent*.

Therefore we also have

$$P(\text{Face Card}|\text{Hearts}) = P(\text{Face Card}) \quad (= \frac{3}{13}\ )\ .$$

**EXERCISE** :

Which of the following pairs of events are independent?

(1) drawing "Hearts" and drawing "Black" ,

(2) drawing "Black" and drawing "Ace" ,

(3) the event $\{2, 3, \cdots, 9\}$ and drawing "Red" .

**EXERCISE** : *Two* numbers are drawn at random from the set

$$\{ 1 , 2 , 3 , 4 \} .$$

If *order is not important* then what is the sample space $\mathcal{S}$ ?

Define the following functions on $\mathcal{S}$ :

$$X( \{i, j\} ) = i + j , \qquad Y( \{i, j\} ) = |i - j| .$$

Which of the following pairs of events are independent?

$$(1) \quad X = 5 \quad \text{and} \quad Y = 2 ,$$

$$(2) \quad X = 5 \quad \text{and} \quad Y = 1 .$$

**REMARK** :

$X$ and $Y$ are examples of *random variables* . (More soon!)

**EXAMPLE** : If $E$ and $F$ are *independent* then so are $E$ and $F^c$ .

**PROOF** : $E = E(F \cup F^c) = EF \cup EF^c$ , where

$EF$ and $EF^c$ are *disjoint* .

Thus

$$P(E) = P(EF) + P(EF^c) ,$$

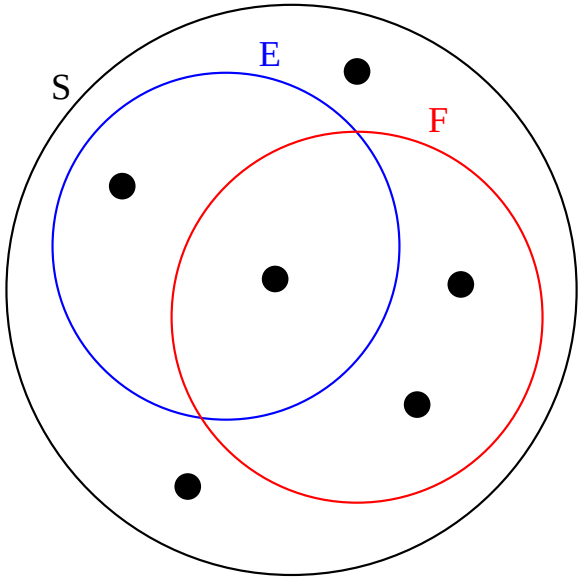
from which

$$\begin{aligned} P(EF^c) &= P(E) - P(EF) \\ &= P(E) - P(E) \cdot P(F) \quad \text{(since } E \text{ and } F \text{ independent)} \\ &= P(E) \cdot ( 1 - P(F) ) \\ &= P(E) \cdot P(F^c) \ . \end{aligned}$$

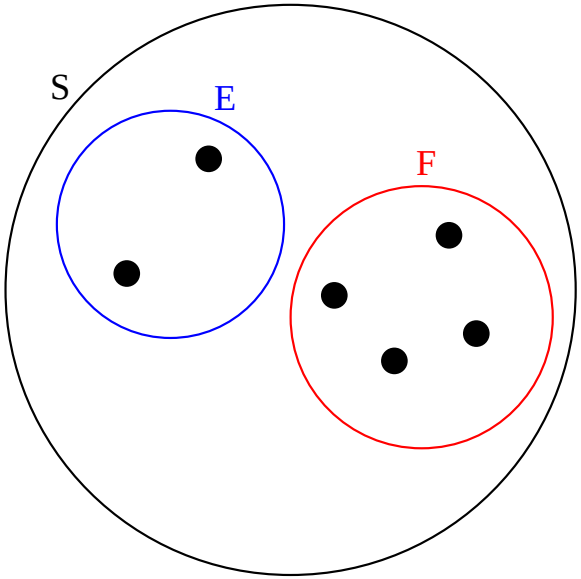
**EXERCISE** :

Prove that if $E$ and $F$ are *independent* then so are $E^c$ and $F^c$ .

**NOTE** : *Independence* and *disjointness* *are different things* !




Independent, but not disjoint.




Disjoint, but not independent.

(The six outcomes in $S$ are assumed to have equal probability.)

If $E$ and $F$ are *independent* then $P(EF) = P(E)\ P(F)$ .

If $E$ and $F$ are *disjoint* then $P(EF) = P(\ \emptyset\ ) = 0$ .

If $E$ and $F$ are *independent and disjoint* then one has *zero probability* !

*Three events* $E$ , $F$, and $G$ are *independent* if

$$P(EFG) \;=\; P(E)\; P(F)\; P(G)\;.$$

*and*

$$\begin{aligned} P(EF) &= P(E)\; P(F)\;. \\ P(EG) &= P(E)\; P(G)\;. \\ P(FG) &= P(F)\; P(G)\;. \end{aligned}$$

**EXERCISE** : Are the three events of drawing

(1) a red card ,

(2) a face card ,

(3) a Heart or Spade ,

independent ?

**EXERCISE** :

A machine $M$ consists of three *independent parts*, $M_1$, $M_2$, and $M_3$ .

Suppose that

$M_1$ functions properly with probability $\frac{9}{10}$ ,

$M_2$ functions properly with probability $\frac{9}{10}$ ,

$M_3$ functions properly with probability $\frac{8}{10}$ ,

and that

the machine $M$ functions if and only if *its three parts function*.

- What is the probability for the machine $M$ to *function* ?
- What is the probability for the machine $M$ to *malfunction* ?

# DISCRETE RANDOM VARIABLES

**DEFINITION** : A *discrete random variable* is a *function* $X(s)$ from a *finite* or *countably infinite* sample space $\mathcal{S}$ to the real numbers :

$$X(\cdot) \quad : \quad \mathcal{S} \quad \rightarrow \quad \mathbb{R} \, .$$

**EXAMPLE** : Toss a coin 3 times in sequence. The sample space is

$$\mathcal{S} \; = \; \{HHH \, , \, HHT \, , \, HTH \, , \, HTT \, , \, THH \, , \, THT \, , \, TTH \, , \, TTT\} \, ,$$

and examples of random variables are

- $X(s) \; =$ the number of Heads in the sequence ; *e.g.*, $X(HTH) \; = \; 2 \, ,$

- $Y(s) \; =$ The index of the first $H$ ; *e.g.*, $Y(TTH) \; = \; 3 \, ,$

  0 if the sequence has no $H$ , *i.e.*, $Y(TTT) \; = \; 0 \, .$

**NOTE** : In this example $X(s)$ and $Y(s)$ are actually *integers* .

*Value-ranges* of a random variable correspond to *events* in $\mathcal{S}$ .

**EXAMPLE** : For the sample space

$$\mathcal{S} = \{HHH\,,\, HHT\,,\, HTH\,,\, HTT\,,\, THH\,,\, THT\,,\, TTH\,,\, TTT\}\,,$$

with

$$X(s) = \text{ the number of Heads }\,,$$

the value

$X(s) = 2\,,$ corresponds to the event $\{HHT\,,\, HTH\,,\, THH\}\,,$

and the values

$1 < X(s) \leq 3\,,$ correspond to $\{HHH\,,\, HHT\,,\, HTH\,,\, THH\}\,.$

**NOTATION** : If it is clear what $\mathcal{S}$ is then we often just write

$$X \qquad \text{instead of} \qquad X(s)\,.$$

*Value-ranges* of a random variable correspond to *events* in $\mathcal{S}$ ,

and

*events* in $\mathcal{S}$ have a *probability* .

Thus

*Value-ranges* of a random variable have a *probability* .

**EXAMPLE** : For the sample space

$$\mathcal{S} = \{HHH\,,\, HHT\,,\, HTH\,,\, HTT\,,\, THH\,,\, THT\,,\, TTH\,,\, TTT\}\,,$$

with

$$X(s) = \text{the number of Heads}\,,$$

we have

$$P(0 < X \leq 2) = \frac{6}{8}\,.$$

**QUESTION** : What are the values of

$$P(X \leq -1)\,,\; P(X \leq 0)\,,\; P(X \leq 1)\,,\; P(X \leq 2)\,,\; P(X \leq 3)\,,\; P(X \leq 4)\,?$$

**NOTATION** : We will also write $p_X(x)$ to denote $P(X = x)$ .

**EXAMPLE** : For the sample space

$$\mathcal{S} = \{HHH\,,\, HHT\,,\, HTH\,,\, HTT\,,\, THH\,,\, THT\,,\, TTH\,,\, TTT\}\,,$$

with

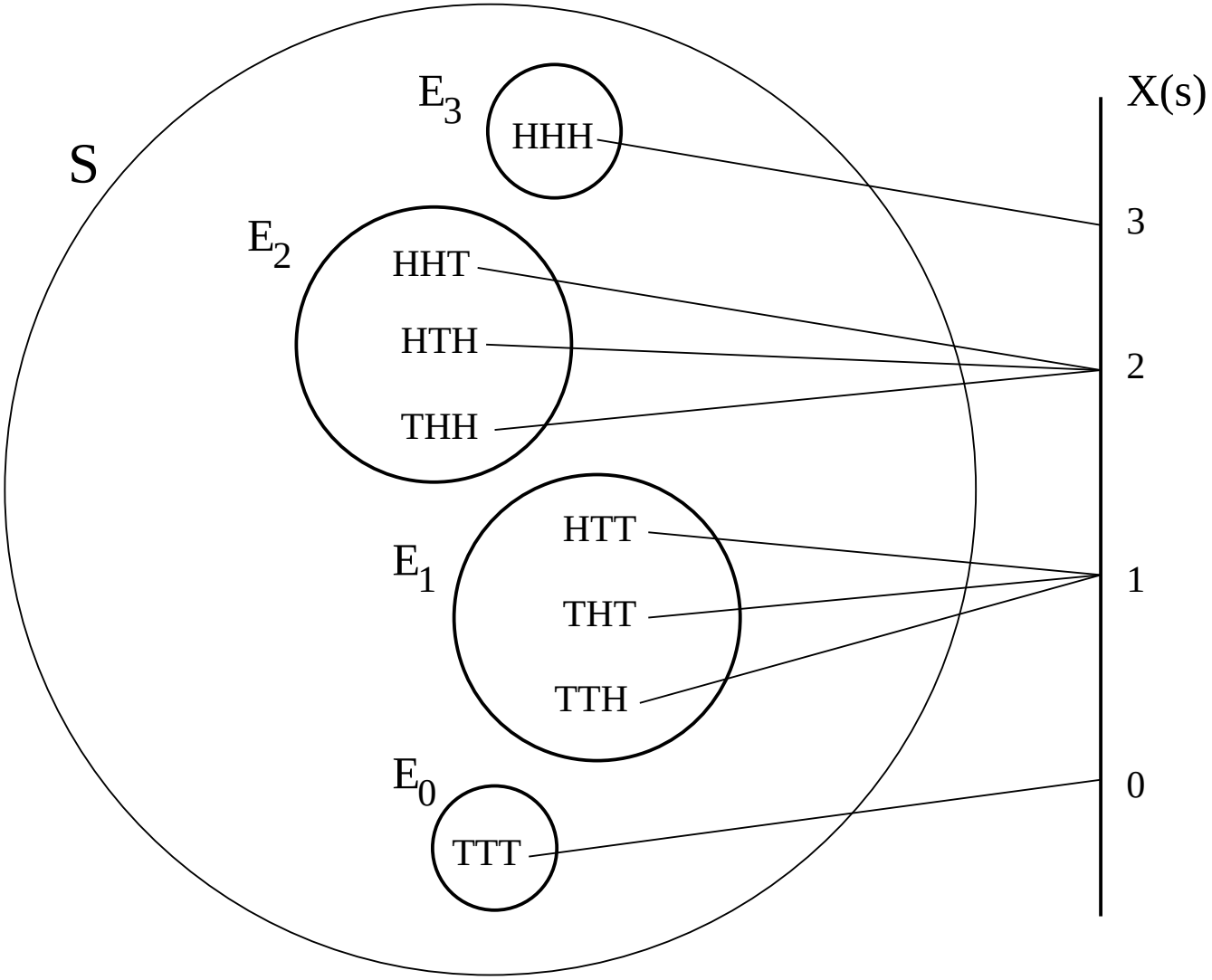
$$X(s) = \text{ the number of Heads }\,,$$

we have

$$\begin{aligned}
p_X(0) &\equiv P(\,\{TTT\}\,) &&= \tfrac{1}{8}\,,\\
p_X(1) &\equiv P(\,\{HTT\,,\, THT\,,\, TTH\}\,) &&= \tfrac{3}{8}\,,\\
p_X(2) &\equiv P(\,\{HHT\,,\, HTH\,,\, THH\}\,) &&= \tfrac{3}{8}\,,\\
p_X(3) &\equiv P(\,\{HHH\}\,) &&= \tfrac{1}{8}\,,
\end{aligned}$$

where

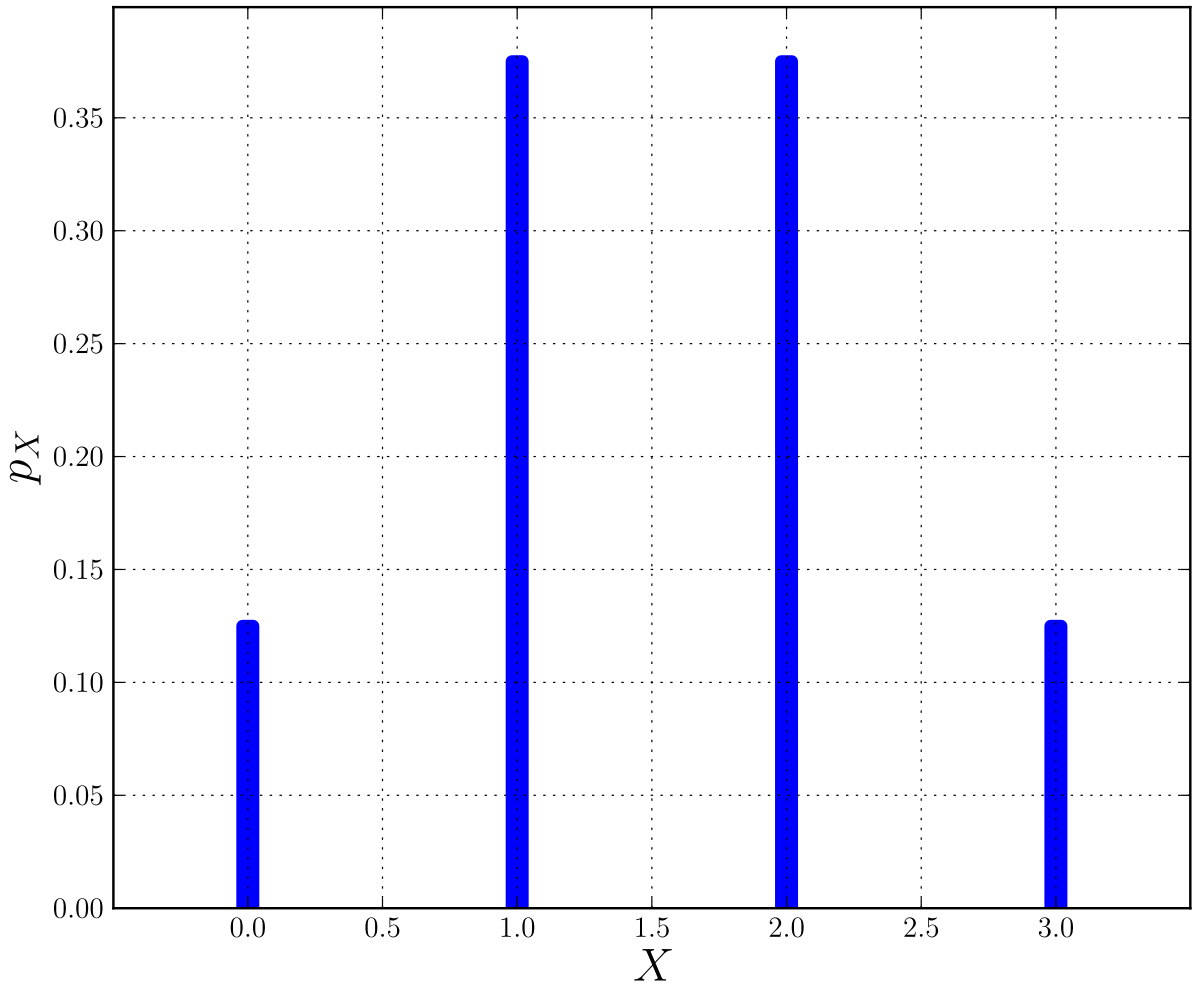
$$p_X(0) + p_X(1) + p_X(2) + p_X(3) = 1\,. \qquad (\text{ Why ? })$$




Graphical representation of $X$ .

The *events* $E_0, E_1, E_2, E_3$ are *disjoint* since $X(s)$ is a *function* !

($X : S \to \mathbb{R}$ must be defined for *all* $s \in S$ and must be *single-valued*.)




The graph of $p_X$ .

**DEFINITION** :

$$p_X(x) \equiv P(X = x) ,$$

is called the *probability mass function* .

**DEFINITION** :

$$F_X(x) \equiv P(X \leq x) ,$$

is called the (*cumulative*) *probability distribution function* .

**PROPERTIES** :

- $F_X(x)$ is a *non-decreasing* function of $x$ . ( Why ? )
- $F_X(-\infty) = 0$ and $F_X(\infty) = 1$ . ( Why ? )
- $P(a < X \leq b) = F_X(b) - F_X(a)$ . ( Why ? )

**NOTATION** : When it is clear what $X$ is then we also write

$$p(x) \text{ for } p_X(x) \qquad \text{and} \qquad F(x) \text{ for } F_X(x) .$$

**EXAMPLE** : With $X(s) =$ the number of Heads , and

$$\mathcal{S} = \{HHH\,,\, HHT\,,\, HTH\,,\, HTT\,,\, THH\,,\, THT\,,\, TTH\,,\, TTT\}\,,$$

$$p(0) = \tfrac{1}{8} \quad , \quad p(1) = \tfrac{3}{8} \quad , \quad p(2) = \tfrac{3}{8} \quad , \quad p(3) = \tfrac{1}{8}\,,$$

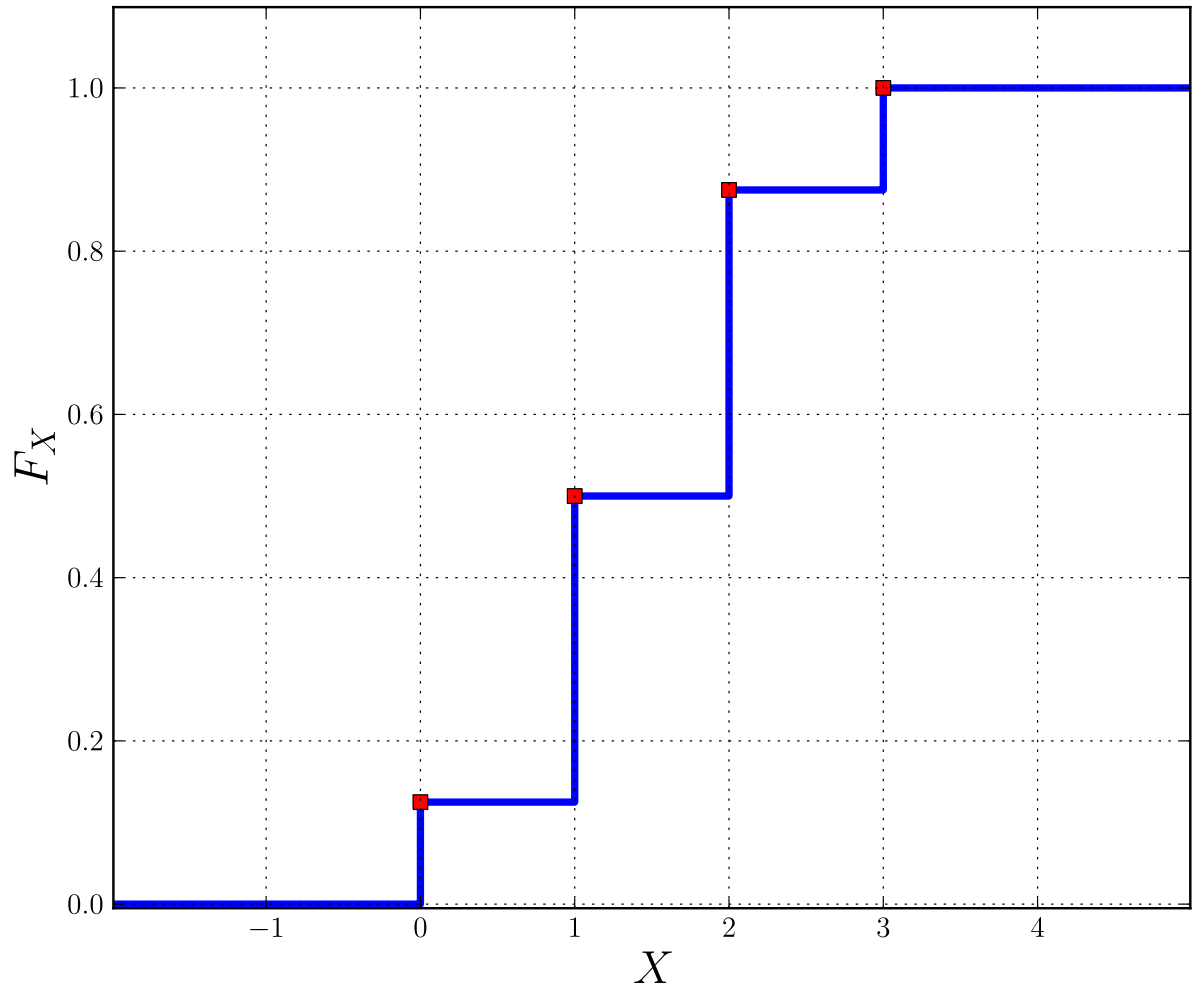
we have the *probability distribution function*

$$\begin{array}{lcll} F(-1) & \equiv & P(X \le -1) & = \quad 0\ , \\ F(\ 0) & \equiv & P(X \le\ 0) & = \quad \mathbf{\frac{1}{8}}\ , \\ F(\ 1) & \equiv & P(X \le\ 1) & = \quad \frac{4}{8}\ , \\ F(\ 2) & \equiv & P(X \le\ 2) & = \quad \mathbf{\frac{7}{8}}\ , \\ F(\ 3) & \equiv & P(X \le\ 3) & = \quad 1\ , \\ F(\ 4) & \equiv & P(X \le\ 4) & = \quad 1\ . \end{array}$$

We see, for example, that

$$P(0 < X \le 2) = P(X = 1) + P(X = 2)$$

$$= F(2) - F(0) = \tfrac{7}{8} - \tfrac{1}{8} = \tfrac{6}{8}\,.$$




The graph of the *probability distribution function* $F_X$ .

**EXAMPLE** : Toss a coin until "Heads" occurs.

Then the sample space is *countably infinite* , namely,

$$\mathcal{S} \;=\; \{H \;,\; TH \;,\; TTH \;,\; TTTH \;,\; \cdots\} \,.$$

The *random variable* $X$ is the *number of rolls* until "Heads" occurs :

$$X(H) \;=\; 1 \quad , \quad X(TH) \;=\; 2 \quad , \quad X(TTH) \;=\; 3 \quad , \; \cdots$$

Then

$$p(1) = \tfrac{1}{2} \quad , \quad p(2) = \tfrac{1}{4} \quad , \quad p(3) = \tfrac{1}{8} \quad , \quad \cdots \quad (\text{ Why ? })$$

and

$$F(n) \;=\; P(X \le n) \;=\; \sum_{k=1}^{n} p(k) \;=\; \sum_{k=1}^{n} \frac{1}{2^k} \;=\; 1 - \frac{1}{2^n} \,,$$

and, as should be the case,

$$\sum_{k=1}^{\infty} p(k) \;=\; \lim_{n\to\infty} \sum_{k=1}^{n} p(k) \;=\; \lim_{n\to\infty} \left(1 - \frac{1}{2^n}\right) \;=\; 1 \,.$$

**NOTE** : The outcomes in $\mathcal{S}$ *do not have equal probability* !

**EXERCISE** : Draw the *probability mass* and *distribution functions*.

$X(s)$ is the *number of tosses* until "Heads" occurs $\cdots$

**REMARK** : We can also take $\mathcal{S} \equiv \mathcal{S}_n$ as *all ordered outcomes of length* $n$. For example, for $n = 4$,

$$\begin{aligned}
\mathcal{S}_4 \;=\; \{ \;\; & \tilde{H}HHH \;, \quad \tilde{H}HHT \;, \quad \tilde{H}HTH \;, \quad \tilde{H}HTT \;, \\
& \tilde{H}THH \;, \quad \tilde{H}THT \;, \quad \tilde{H}TTH \;, \quad \tilde{H}TTT \;, \\
& T\tilde{H}HH \;, \quad T\tilde{H}HT \;, \quad T\tilde{H}TH \;, \quad T\tilde{H}TT \;, \\
& TT\tilde{H}H \;, \quad TT\tilde{H}T \;, \quad TTT\tilde{H} \;, \quad TTTT \quad \} \;.
\end{aligned}$$

where in each outcome the first "Heads" is marked as $\tilde{H}$ .

Each outcome in $\mathcal{S}_4$ has *equal probability* $2^{-n}$ (here $2^{-4} = \frac{1}{16}$) , and

$$p_X(1) = \tfrac{1}{2} \quad , \quad p_X(2) = \tfrac{1}{4} \quad , \quad p_X(3) = \tfrac{1}{8} \quad , \quad p_X(4) = \tfrac{1}{16} \quad \cdots \;,$$

independent of $n$ .

## Joint distributions

The *probability mass function* and the *probability distribution function* can also be functions of *more than one variable*.

**EXAMPLE** : Toss a coin 3 times in sequence. For the sample space

$$\mathcal{S} = \{HHH\,,\, HHT\,,\, HTH\,,\, HTT\,,\, THH\,,\, THT\,,\, TTH\,,\, TTT\}\,,$$

we let

$$X(s) = \#\text{ Heads} \quad , \quad Y(s) = \text{index of the first } H \quad (0 \text{ for } TTT)\,.$$

Then we have the *joint probability mass function*

$$p_{X,Y}(x,y) = P(X = x\ ,\ Y = y)\,.$$

For example,

$$\begin{aligned} p_{X,Y}(2,1) &= P(X = 2\ ,\ Y = 1) \\ &= P(\text{ 2 Heads } ,\ 1^{\text{st}} \text{ toss is Heads}) \\ &= \tfrac{2}{8} = \tfrac{1}{4}\,. \end{aligned}$$

**EXAMPLE** : ( continued $\cdots$ ) For

$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$,

$X(s) =$ number of Heads, and $Y(s) =$ index of the first $H$ ,

we can list the values of $p_{X,Y}(x, y)$ :

Joint probability mass function $p_{X,Y}(x, y)$

| | $y = 0$ | $\mathbf{y = 1}$ | $y = 2$ | $y = 3$ | $p_X(x)$ |
|---|---|---|---|---|---|
| $x = 0$ | $\frac{1}{8}$ | 0 | 0 | 0 | $\frac{1}{8}$ |
| $x = 1$ | 0 | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{3}{8}$ |
| $\mathbf{x = 2}$ | 0 | $\frac{\mathbf{2}}{\mathbf{8}}$ | $\frac{1}{8}$ | 0 | $\frac{3}{8}$ |
| $x = 3$ | 0 | $\frac{1}{8}$ | 0 | 0 | $\frac{1}{8}$ |
| $p_Y(y)$ | $\frac{1}{8}$ | $\frac{4}{8}$ | $\frac{2}{8}$ | $\frac{1}{8}$ | 1 |

**NOTE** :

- The *marginal probability* $p_X$ is the probability mass function of $X$.
- The *marginal probability* $p_Y$ is the probability mass function of $Y$.

**EXAMPLE** : ( continued $\cdots$ )

$X(s)$ = number of Heads, and $Y(s)$ = index of the first $H$ .

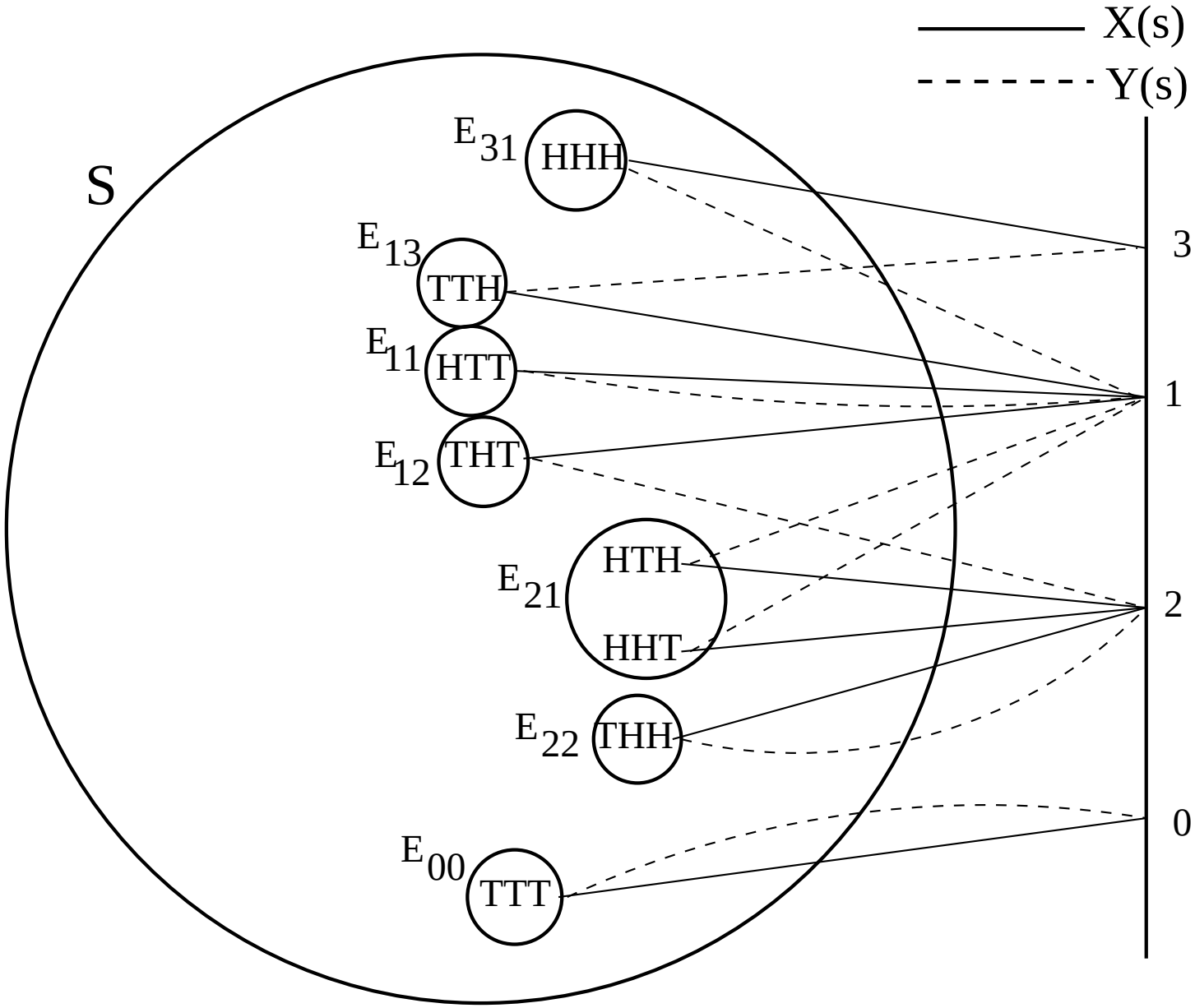
| | $y=0$ | $\mathbf{y=1}$ | $y=2$ | $y=3$ | $p_X(x)$ |
|---|---|---|---|---|---|
| $x=0$ | $\frac{1}{8}$ | 0 | 0 | 0 | $\frac{1}{8}$ |
| $x=1$ | 0 | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{3}{8}$ |
| $\mathbf{x=2}$ | 0 | $\frac{\mathbf{2}}{\mathbf{8}}$ | $\frac{1}{8}$ | 0 | $\frac{\mathbf{3}}{\mathbf{8}}$ |
| $x=3$ | 0 | $\frac{1}{8}$ | 0 | 0 | $\frac{1}{8}$ |
| $p_Y(y)$ | $\frac{1}{8}$ | $\frac{\mathbf{4}}{\mathbf{8}}$ | $\frac{2}{8}$ | $\frac{1}{8}$ | 1 |

For example,

- $X=2$ corresponds to the *event* $\{HHT\ ,\ HTH\ ,\ THH\}$ .
- $Y=1$ corresponds to the *event* $\{HHH\ ,\ HHT\ ,\ HTH\ ,\ HTT\}$ .
- $(X=2$ *and* $Y=1)$ corresponds to the *event* $\{HHT\ ,\ HTH\}$ .

**QUESTION** :

Are the event $X=2$ and the event $Y=1$ *independent* ?




The *events* $E_{i,j} \equiv \{\ s \in S\ :\ X(s) = i\ ,\ Y(s) = j\ \}$ are *disjoint* .

**DEFINITION** :

$$p_{X,Y}(x,y) \quad \equiv \quad P(X = x \ , \ Y = y) \ ,$$

is called the *joint probability mass function* .

**DEFINITION** :

$$F_{X,Y}(x,y) \quad \equiv \quad P(X \leq x \ , \ Y \leq y) \ ,$$

is called the *joint* (*cumulative*) *probability distribution function* .

**NOTATION** : When it is clear what $X$ and $Y$ are then we also write

$$p(x,y) \quad \text{for} \quad p_{X,Y}(x,y) \ ,$$

and

$$F(x,y) \quad \text{for} \quad F_{X,Y}(x,y) \ .$$

**EXAMPLE** : *Three* tosses : $X(s) = \#$ Heads, $Y(s) =$ index $1^{\text{st}}$ $H$ .

*Joint probability mass function* $p_{X,Y}(x, y)$

| | $y=0$ | $y=1$ | $y=2$ | $y=3$ | $p_X(x)$ |
|---|---|---|---|---|---|
| $x=0$ | $\frac{1}{8}$ | $0$ | $0$ | $0$ | $\frac{1}{8}$ |
| $x=1$ | $0$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{3}{8}$ |
| $x=2$ | $0$ | $\frac{2}{8}$ | $\frac{1}{8}$ | $0$ | $\frac{3}{8}$ |
| $x=3$ | $0$ | $\frac{1}{8}$ | $0$ | $0$ | $\frac{1}{8}$ |
| $p_Y(y)$ | $\frac{1}{8}$ | $\frac{4}{8}$ | $\frac{2}{8}$ | $\frac{1}{8}$ | $1$ |

*Joint distribution function* $F_{X,Y}(x, y) \equiv P(X \le x, Y \le y)$

| | $y=0$ | $y=1$ | $y=2$ | $y=3$ | $F_X(\cdot)$ |
|---|---|---|---|---|---|
| $x=0$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ |
| $x=1$ | $\frac{1}{8}$ | $\frac{2}{8}$ | $\frac{3}{8}$ | $\frac{4}{8}$ | $\frac{4}{8}$ |
| $x=2$ | $\frac{1}{8}$ | $\frac{4}{8}$ | $\frac{6}{8}$ | $\frac{7}{8}$ | $\frac{7}{8}$ |
| $x=3$ | $\frac{1}{8}$ | $\frac{5}{8}$ | $\frac{7}{8}$ | $1$ | $1$ |
| $F_Y(\cdot)$ | $\frac{1}{8}$ | $\frac{5}{8}$ | $\frac{7}{8}$ | $1$ | $1$ |

Note that the distribution function $F_X$ is a *copy* of the 4th column, and the distribution function $F_Y$ is a *copy* of the 4th row. ( Why ? )

In the preceding example :

Joint probability mass function $p_{X,Y}(x,y)$

| | $y=0$ | $y=1$ | $y=2$ | $y=3$ | $p_X(x)$ |
|---|---|---|---|---|---|
| $x=0$ | $\frac{1}{8}$ | 0 | 0 | 0 | $\frac{1}{8}$ |
| $x=1$ | 0 | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{3}{8}$ |
| $x=2$ | 0 | $\frac{2}{8}$ | $\frac{1}{8}$ | **0** | $\frac{3}{8}$ |
| $x=3$ | 0 | $\frac{1}{8}$ | **0** | **0** | $\frac{1}{8}$ |
| $p_Y(y)$ | $\frac{1}{8}$ | $\frac{4}{8}$ | $\frac{2}{8}$ | $\frac{1}{8}$ | 1 |

*Joint distribution function* $F_{X,Y}(x,y) \equiv P(X \le x, Y \le y)$

| | $y=0$ | $y=1$ | $y=2$ | $y=3$ | $F_X(\cdot)$ |
|---|---|---|---|---|---|
| $x=0$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ |
| $x=1$ | $\frac{1}{8}$ | $\frac{2}{8}$ | $\frac{3}{8}$ | $\frac{4}{8}$ | $\frac{4}{8}$ |
| $x=2$ | $\frac{1}{8}$ | $\frac{4}{8}$ | $\frac{6}{8}$ | $\frac{7}{8}$ | $\frac{7}{8}$ |
| $x=3$ | $\frac{1}{8}$ | $\frac{5}{8}$ | $\frac{7}{8}$ | **1** | 1 |
| $F_Y(\cdot)$ | $\frac{1}{8}$ | $\frac{5}{8}$ | $\frac{7}{8}$ | 1 | 1 |

**QUESTION** : Why is

$$P(1 < X \le 3\,, 1 < Y \le 3) \;=\; F(3,3) - F(1,3) - F(3,1) + F(1,1)\,?$$

**EXERCISE** :

Roll a *four-sided die* (tetrahedron) *two* times.

(The sides are marked 1 , 2 , 3 , 4 .)

Suppose each of the four sides is equally likely to end facing down.

Suppose the *outcome* of a *single roll* is the side that faces *down* (!).

Define the random variables $X$ and $Y$ as

$X =$ result of the *first roll* , $Y =$ *sum* of the two rolls.

- What is a good choice of the *sample space* $\mathcal{S}$ ?
- How many outcomes are there in $\mathcal{S}$ ?
- List the values of the *joint probability mass function* $p_{X,Y}(x, y)$ .
- List the values of the *joint cumulative distribution function* $F_{X,Y}(x, y)$ .

**EXERCISE** :

*Three* balls are selected at random from a bag containing

$$2 \ \textit{red} \ , \ 3 \ \textit{green} \ , \ 4 \ \textit{blue} \ \text{balls} \ .$$

Define the *random variables*

$$R(s) \ = \ \text{the number of } \textit{red} \text{ balls drawn,}$$

and

$$G(s) \ = \ \text{the number of } \textit{green} \text{ balls drawn} \ .$$

List the values of

- the *joint probability mass function* $p_{R,G}(r, g)$ .
- the *marginal probability mass functions* $p_R(r)$ and $p_G(g)$ .
- the *joint distribution function* $F_{R,G}(r, g)$ .
- the *marginal distribution functions* $F_R(r)$ and $F_G(g)$ .

## Independent random variables

*Two* discrete random variables $X(s)$ and $Y(s)$ are *independent* if

$$P(X = x \ , \ Y = y) \ = \ P(X = x) \cdot P(Y = y) \ , \quad \text{for } \textit{all } x \text{ and } y \ ,$$

or, equivalently, if their *probability mass functions* satisfy

$$p_{X,Y}(x, y) \ = \ p_X(x) \ \cdot \ p_Y(y) \ , \quad \text{for } \textit{all } x \text{ and } y \ ,$$

or, equivalently, if the *events*

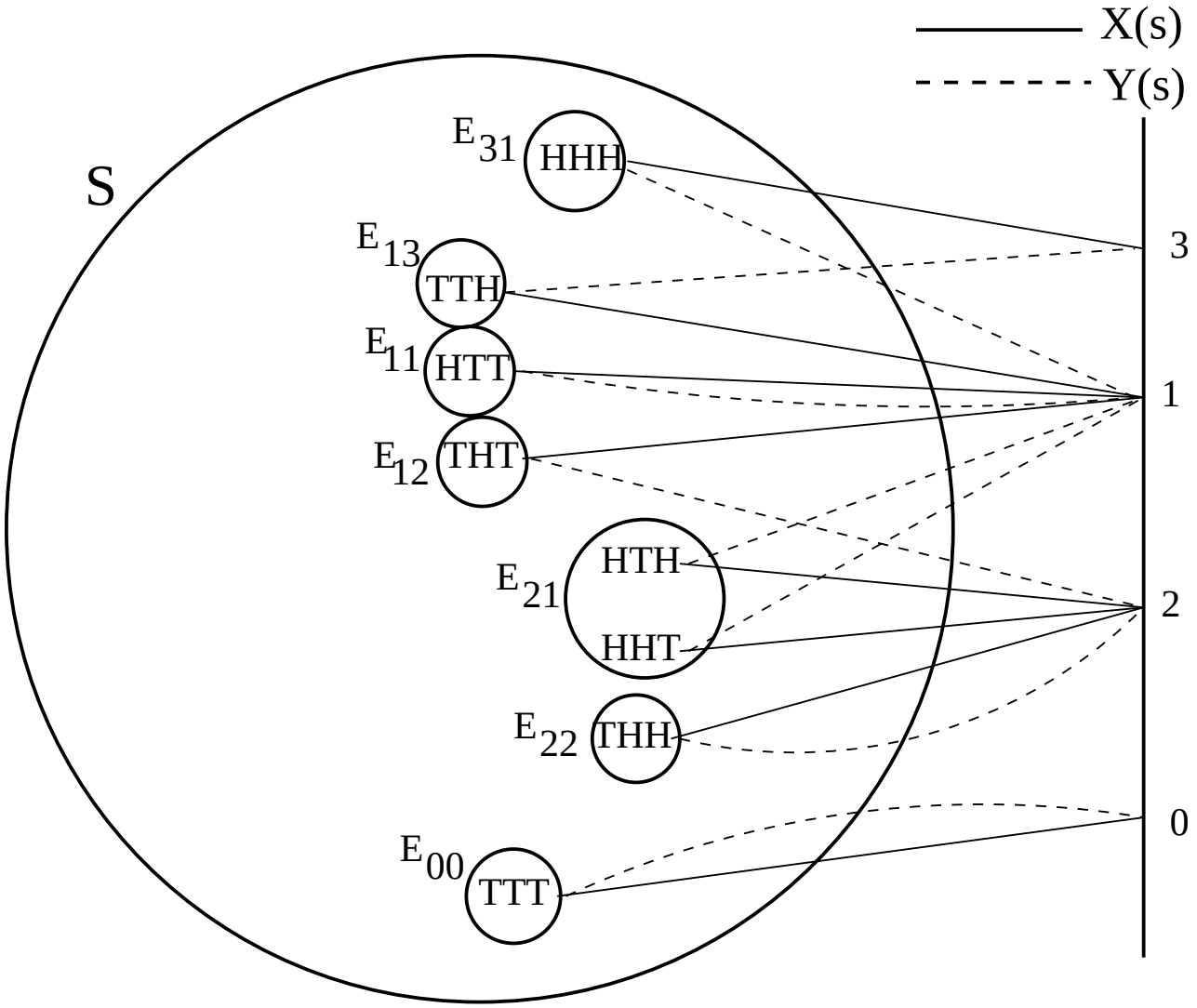
$$E_x \ \equiv \ X^{-1}(\{x\}) \quad \text{and} \quad E_y \ \equiv \ Y^{-1}(\{y\}) \ ,$$

are independent *in the sample space* $\mathcal{S}$ , *i.e.*,

$$P(E_x E_y) \ = \ P(E_x) \cdot P(E_y) \ , \quad \text{for } \textit{all } x \text{ and } y \ .$$

**NOTE** :

- In the current *discrete* case, $x$ and $y$ are typically *integers* .
- $X^{-1}(\{x\}) \ \equiv \ \{ \ s \in \mathcal{S} \ : \ X(s) = x \ \} \ .$




*Three* tosses : $X(s) = \#$ Heads, $Y(s) =$ index $1^{\text{st}}$ $H$ .

- What are the values of $p_X(2)$ , $p_Y(1)$ , $p_{X,Y}(2,1)$ ?
- Are $X$ and $Y$ *independent* ?

**RECALL** :

$X(s)$ and $Y(s)$ are *independent* if for all $x$ and $y$ :

$$p_{X,Y}(x,y) \;=\; p_X(x) \cdot p_Y(y) \; .$$

**EXERCISE** :

*Roll a die two times* in a row.

Let

$$X \text{ be the result of the } 1^{\text{st}} \text{ roll} \; ,$$

and

$$Y \text{ the result of the } 2^{\text{nd}} \text{ roll} \; .$$

Are $X$ and $Y$ *independent* , *i.e.*, is

$$p_{X,Y}(k,\ell) \;=\; p_X(k) \cdot p_Y(\ell), \qquad \text{for all } \; 1 \;\leq\; k,\ell \;\leq\; 6 \;\; ?$$

**EXERCISE** :

Are these random variables $X$ and $Y$ *independent* ?

Joint probability mass function $p_{X,Y}(x,y)$

| | $y=0$ | $y=1$ | $y=2$ | $y=3$ | $p_X(x)$ |
|---|---|---|---|---|---|
| $x=0$ | $\frac{1}{8}$ | $0$ | $0$ | $0$ | $\frac{1}{8}$ |
| $x=1$ | $0$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{3}{8}$ |
| $x=2$ | $0$ | $\frac{2}{8}$ | $\frac{1}{8}$ | $0$ | $\frac{3}{8}$ |
| $x=3$ | $0$ | $\frac{1}{8}$ | $0$ | $0$ | $\frac{1}{8}$ |
| $p_Y(y)$ | $\frac{1}{8}$ | $\frac{4}{8}$ | $\frac{2}{8}$ | $\frac{1}{8}$ | $1$ |

**EXERCISE** : Are these random variables $X$ and $Y$ *independent* ?

Joint probability mass function $p_{X,Y}(x,y)$

| | $y=1$ | $y=2$ | $y=3$ | $p_X(x)$ |
|---|---|---|---|---|
| $x=1$ | $\frac{1}{3}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{2}$ |
| $x=2$ | $\frac{2}{9}$ | $\frac{1}{18}$ | $\frac{1}{18}$ | $\frac{1}{3}$ |
| $x=3$ | $\frac{1}{9}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{6}$ |
| $p_Y(y)$ | $\frac{2}{3}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | 1 |

Joint distribution function $F_{X,Y}(x,y) \equiv P(X \leq x, Y \leq y)$

| | $y=1$ | $y=2$ | $y=3$ | $F_X(x)$ |
|---|---|---|---|---|
| $x=1$ | $\frac{1}{3}$ | $\frac{5}{12}$ | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $x=2$ | $\frac{5}{9}$ | $\frac{25}{36}$ | $\frac{5}{6}$ | $\frac{5}{6}$ |
| $x=3$ | $\frac{2}{3}$ | $\frac{5}{6}$ | 1 | 1 |
| $F_Y(y)$ | $\frac{2}{3}$ | $\frac{5}{6}$ | 1 | 1 |

**QUESTION** : Is $F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$ ?

**PROPERTY** :

The *joint distribution function* of *independent* random variables $X$ and $Y$ satisfies

$$F_{X,Y}(x,y) \quad = \quad F_X(x) \cdot F_Y(y) \ , \qquad \text{for all } x, y \ .$$

**PROOF** :

$$\begin{aligned}
F_{X,Y}(x_k, y_\ell) \quad &= \quad P(X \leq x_k \ , \ Y \leq y_\ell) \\
&= \quad \sum_{i \leq k} \ \sum_{j \leq \ell} \ p_{X,Y}(x_i, y_j) \\
&= \quad \sum_{i \leq k} \ \sum_{j \leq \ell} \ p_X(x_i) \cdot p_Y(y_j) \qquad \text{(by independence)} \\
&= \quad \sum_{i \leq k} \ \{ \ p_X(x_i) \cdot \sum_{j \leq \ell} \ p_Y(y_j) \ \} \\
&= \quad \{ \sum_{i \leq k} \ p_X(x_i) \ \} \ \cdot \ \{ \sum_{j \leq \ell} \ p_Y(y_j) \ \} \\
&= \quad F_X(x_k) \cdot F_Y(y_\ell) \ .
\end{aligned}$$

## Conditional distributions

Let $X$ and $Y$ be discrete random variables with *joint probability mass function*

$$p_{X,Y}(x, y) \ .$$

For given $x$ and $y$ , let

$$E_x \ = \ X^{-1}(\{x\}) \qquad \text{and} \qquad E_y \ = \ Y^{-1}(\{y\}) \ ,$$

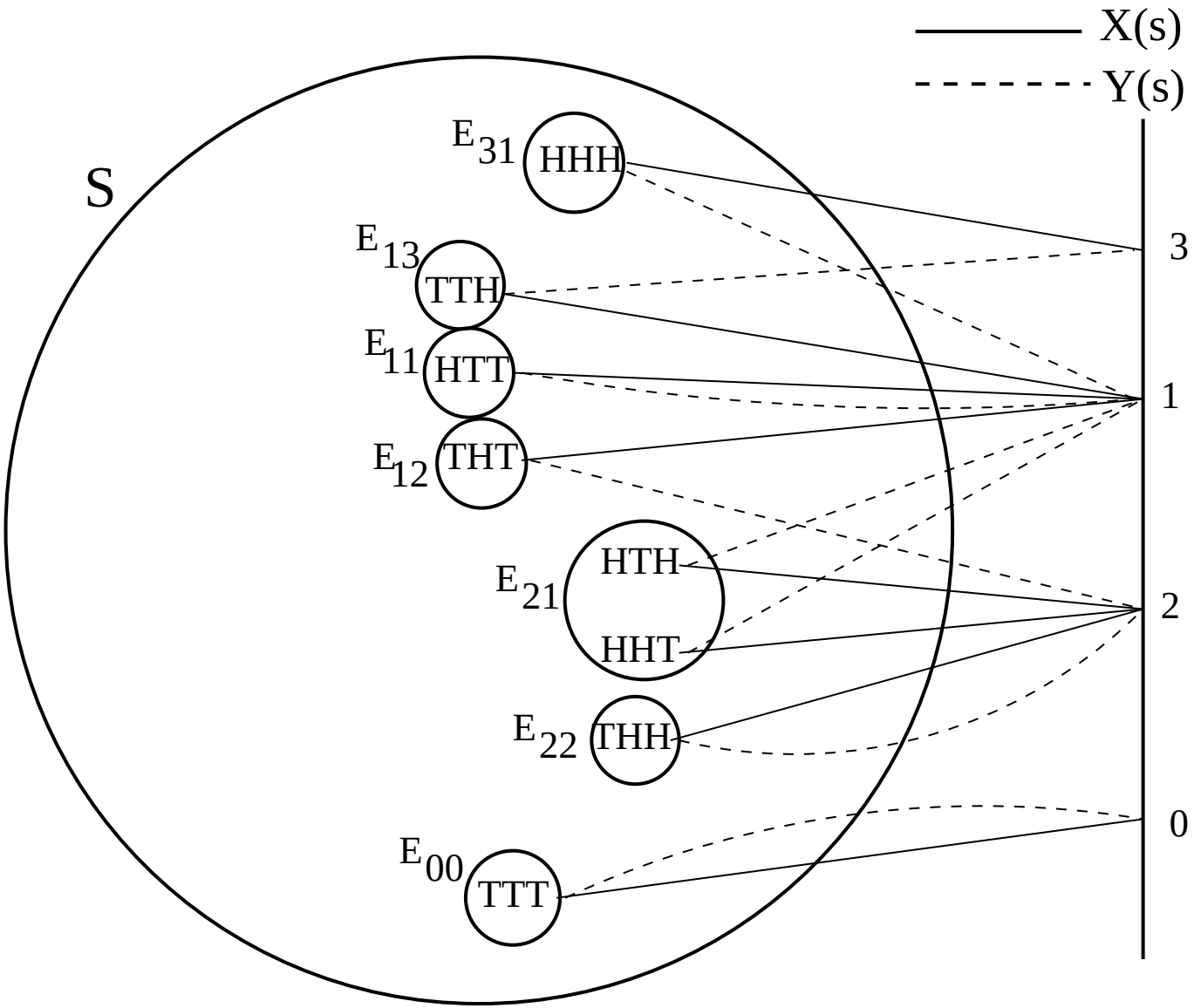
be their corresponding *events* in the sample space $\mathcal{S}$.

Then

$$P(E_x|E_y) \ \equiv \ \frac{P(E_x E_y)}{P(E_y)} \ = \ \frac{p_{X,Y}(x, y)}{p_Y(y)} \ .$$

Thus it is natural to define the *conditional probability mass function*

$$p_{X|Y}(x|y) \ \equiv \ P(X = x \mid Y = y) \ = \ \frac{p_{X,Y}(x, y)}{p_Y(y)} \ .$$




*Three* tosses : $X(s) = \#$ Heads, $Y(s) =$ index $1^{\text{st}}$ $H$ .

- What are the values of $P(X = 2 \,|\, Y = 1)$ and $P(Y = 1 \,|\, X = 2)$ ?

**EXAMPLE** : (3 tosses : $X(s) = \#$ Heads, $Y(s) =$ index $1^{\text{st}}$ $H$.)

Joint probability mass function $p_{X,Y}(x, y)$

| | $y=0$ | $y=1$ | $y=2$ | $y=3$ | $p_X(x)$ |
|---|---|---|---|---|---|
| $x=0$ | $\frac{1}{8}$ | 0 | 0 | 0 | $\frac{1}{8}$ |
| $x=1$ | 0 | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{3}{8}$ |
| $x=2$ | 0 | $\frac{2}{8}$ | $\frac{1}{8}$ | 0 | $\frac{3}{8}$ |
| $x=3$ | 0 | $\frac{1}{8}$ | 0 | 0 | $\frac{1}{8}$ |
| $p_Y(y)$ | $\frac{1}{8}$ | $\frac{4}{8}$ | $\frac{2}{8}$ | $\frac{1}{8}$ | 1 |

Conditional probability mass function $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$ .

| | $y=0$ | $y=1$ | $y=2$ | $y=3$ |
|---|---|---|---|---|
| $x=0$ | 1 | 0 | 0 | 0 |
| $x=1$ | 0 | $\frac{2}{8}$ | $\frac{4}{8}$ | 1 |
| $x=2$ | 0 | $\frac{4}{8}$ | $\frac{4}{8}$ | 0 |
| $x=3$ | 0 | $\frac{2}{8}$ | 0 | 0 |
| | 1 | 1 | 1 | 1 |

**EXERCISE** : Also construct the Table for $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$ .

**EXAMPLE** :

Joint probability mass function $p_{X,Y}(x,y)$

| | $y=1$ | $y=2$ | $y=3$ | $p_X(x)$ |
|---|---|---|---|---|
| $x=1$ | $\frac{1}{3}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{2}$ |
| $x=2$ | $\frac{2}{9}$ | $\frac{1}{18}$ | $\frac{1}{18}$ | $\frac{1}{3}$ |
| $x=3$ | $\frac{1}{9}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{6}$ |
| $p_Y(y)$ | $\frac{2}{3}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | 1 |

Conditional probability mass function $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$ .

| | $y=1$ | $y=2$ | $y=3$ |
|---|---|---|---|
| $x=1$ | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ |
| $x=2$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ |
| $x=3$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |
| | 1 | 1 | 1 |

**QUESTION** : What does the last Table tell us?

**EXERCISE** : Also construct the Table for $P(Y=y|X=x)$ .

## Expectation

The *expected value* of a discrete random variable $X$ is

$$E[X] \equiv \sum_k x_k \cdot P(X = x_k) = \sum_k x_k \cdot p_X(x_k) \ .$$

Thus $E[X]$ represents the *weighted average value* of $X$ .

( $E[X]$ is also called the *mean* of $X$ .)

**EXAMPLE** : The *expected value* of *rolling a die* is

$$E[X] = 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + \cdots + 6 \cdot \tfrac{1}{6} = \tfrac{1}{6} \cdot \textstyle\sum_{k=1}^{6} k = \tfrac{7}{2} \ .$$

**EXERCISE** : Prove the following :

- $E[aX] = a\ E[X]$ ,
- $E[aX + b] = a\ E[X] + b$ .

**EXAMPLE** : Toss a coin until "Heads" occurs. Then

$$\mathcal{S} \;=\; \{H \;,\; TH \;,\; TTH \;,\; TTTH \;,\; \cdots\} \,.$$

The *random variable* $X$ is the *number of tosses* until "Heads" occurs :

$$X(H) \;=\; 1 \quad , \quad X(TH) \;=\; 2 \quad , \quad X(TTH) \;=\; 3 \,.$$

Then

$$E[X] \;=\; 1 \cdot \frac{1}{2} \;+\; 2 \cdot \frac{1}{4} \;+\; 3 \cdot \frac{1}{8} \;+\; \cdots \;=\; \lim_{n \to \infty} \sum_{k=1}^{n} \frac{k}{2^k} \;=\; 2 \,.$$

| $n$ | $\sum_{k=1}^{n} \; k/2^k$ |
|---|---|
| 1 | 0.50000000 |
| 2 | 1.00000000 |
| 3 | 1.37500000 |
| 10 | 1.98828125 |
| 40 | 1.99999999 |

**REMARK** :

Perhaps using $\mathcal{S}_n = \{\text{all sequences of } n \text{ tosses}\}$ is better $\cdots$

The expected value of a *function of a random variable* is

$$E[g(X)] \equiv \sum_k g(x_k)\, p(x_k) \,.$$

**EXAMPLE** :

The *pay-off* of rolling a die is $\$k^2$ , where $k$ is the side facing up.

What should the *entry fee* be for the betting to break even?

**SOLUTION** : Here $g(X) = X^2$ , and

$$E[g(X)] = \sum_{k=1}^{6} k^2 \frac{1}{6} = \frac{1}{6} \frac{6(6+1)(2\cdot 6+1)}{6} = \frac{91}{6} \cong \$15.17 \,.$$

The expected value of a function of *two* random variables is

$$E[g(X,Y)] \equiv \sum_k \sum_\ell g(x_k, y_\ell)\, p(x_k, y_\ell) \ .$$

**EXAMPLE**:

| | $y=1$ | $y=2$ | $y=3$ | $p_X(x)$ |
|---|---|---|---|---|
| $x=1$ | $\frac{1}{3}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{2}$ |
| $x=2$ | $\frac{2}{9}$ | $\frac{1}{18}$ | $\frac{1}{18}$ | $\frac{1}{3}$ |
| $x=3$ | $\frac{1}{9}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{6}$ |
| $p_Y(y)$ | $\frac{2}{3}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | 1 |

$$E[X] = 1\cdot\tfrac{1}{2} + 2\cdot\tfrac{1}{3} + 3\cdot\tfrac{1}{6} = \tfrac{5}{3}\ ,$$

$$E[Y] = 1\cdot\tfrac{2}{3} + 2\cdot\tfrac{1}{6} + 3\cdot\tfrac{1}{6} = \tfrac{3}{2}\ ,$$

$$E[XY] = 1\cdot\tfrac{1}{3} + 2\cdot\tfrac{1}{12} + 3\cdot\tfrac{1}{12}$$

$$+ \ 2\cdot\tfrac{2}{9} + 4\cdot\tfrac{1}{18} + 6\cdot\tfrac{1}{18}$$

$$+ \ 3\cdot\tfrac{1}{9} + 6\cdot\tfrac{1}{36} + 9\cdot\tfrac{1}{36} = \tfrac{5}{2}\ . \qquad (\text{So ?})$$

**PROPERTY** :

- If $X$ and $Y$ are *independent* then $E[XY] = E[X]\,E[Y]$ .

**PROOF** :

$$
\begin{aligned}
E[XY] &= \textstyle\sum_k \sum_\ell \; x_k \; y_\ell \; p_{X,Y}(x_k, y_\ell) \\
&= \textstyle\sum_k \sum_\ell \; x_k \; y_\ell \; p_X(x_k) \; p_Y(y_\ell) \qquad \text{(by independence)} \\
&= \textstyle\sum_k \{ \; x_k \; p_X(x_k) \sum_\ell \; y_\ell \; p_Y(y_\ell)\} \\
&= \{\textstyle\sum_k \; x_k \; p_X(x_k)\} \cdot \{\sum_\ell \; y_\ell \; p_Y(y_\ell)\} \\
&= E[X] \cdot E[Y] \; .
\end{aligned}
$$

**EXAMPLE** : See the preceding example !

**PROPERTY** : $E[X+Y] \;=\; E[X] \;+\; E[Y]$ .

**PROOF** :

$$
\begin{aligned}
E[X+Y] &= \sum_k \sum_\ell (x_k + y_\ell)\, p_{X,Y}(x_k, y_\ell) \\
&= \sum_k \sum_\ell x_k\, p_{X,Y}(x_k, y_\ell) \;+\; \sum_k \sum_\ell y_\ell\, p_{X,Y}(x_k, y_\ell) \\
&= \sum_k \sum_\ell x_k\, p_{X,Y}(x_k, y_\ell) \;+\; \sum_\ell \sum_k y_\ell\, p_{X,Y}(x_k, y_\ell) \\
&= \sum_k \{x_k \sum_\ell p_{X,Y}(x_k, y_\ell)\} \;+\; \sum_\ell \{ y_\ell \sum_k p_{X,Y}(x_k, y_\ell)\} \\
&= \sum_k \{x_k\, p_X(x_k)\} \;+\; \sum_\ell \{y_\ell\, p_Y(y_\ell)\} \\
&= E[X] \;+\; E[Y] \,.
\end{aligned}
$$

**NOTE** : $X$ and $Y$ need not be independent !

**EXERCISE** :

Probability mass function $p_{X,Y}(x,y)$

| | $y=6$ | $y=8$ | $y=10$ | $p_X(x)$ |
|---|---|---|---|---|
| $x=1$ | $\frac{1}{5}$ | $0$ | $\frac{1}{5}$ | $\frac{2}{5}$ |
| $x=2$ | $0$ | $\frac{1}{5}$ | $0$ | $\frac{1}{5}$ |
| $x=3$ | $\frac{1}{5}$ | $0$ | $\frac{1}{5}$ | $\frac{2}{5}$ |
| $p_Y(y)$ | $\frac{2}{5}$ | $\frac{1}{5}$ | $\frac{2}{5}$ | $1$ |

Show that

- $E[X] = 2 \quad , \quad E[Y] = 8 \quad , \quad E[XY] = 16$
- $X$ and $Y$ are *not* independent

Thus if

$$E[XY] = E[X]\, E[Y] \, ,$$

then it does not necessarily follow that $X$ and $Y$ are independent !

## Variance and Standard Deviation

Let $X$ have *mean*

$$\mu \;=\; E[X]\;.$$

Then the *variance* of $X$ is

$$Var(X) \;\equiv\; E[\,(X-\mu)^2] \;\equiv\; \sum_k\,(x_k-\mu)^2\,p(x_k)\;,$$

which is the average weighted *square distance* from the mean.

We have

$$\begin{aligned} Var(X) \;&=\; E[X^2 \;-\; 2\mu X \;+\; \mu^2] \\ &=\; E[X^2] \;-\; 2\mu E[X] \;+\; \mu^2 \\ &=\; E[X^2] \;-\; 2\mu^2 \;+\; \mu^2 \\ &=\; E[X^2] \;-\; \mu^2\;. \end{aligned}$$

The *standard deviation* of $X$ is

$$\sigma(X) \equiv \sqrt{Var(X)} = \sqrt{E[\,(X-\mu)^2]} = \sqrt{E[X^2] - \mu^2}\,.$$

which is the average weighted *distance* from the mean.

**EXAMPLE** : The *variance* of *rolling a die* is

$$Var(X) = \sum_{k=1}^{6} [k^2 \cdot \frac{1}{6}\,] - \mu^2$$

$$= \frac{1}{6}\,\frac{6(6+1)(2\cdot 6+1)}{6} - (\frac{7}{2})^2 = \frac{35}{12}\,.$$

The *standard deviation* is

$$\sigma = \sqrt{\frac{35}{12}} \cong 1.70\,.$$

## Covariance

Let $X$ and $Y$ be random variables with *mean*

$$E[X] \;=\; \mu_X \qquad , \qquad E[Y] \;=\; \mu_Y \;.$$

Then the *covariance* of $X$ and $Y$ is defined as

$$Cov(X,Y) \;\equiv\; E[\,(X-\mu_X)\,(Y-\mu_Y)\,] \;=\; \sum_{k,\ell} (x_k-\mu_X)\,(y_\ell-\mu_Y)\,p(x_k,y_\ell)\;.$$

We have

$$\begin{aligned} Cov(X,Y) \;&=\; E[\,(X-\mu_X)\,(Y-\mu_Y)\,] \\ &=\; E[XY-\mu_X Y-\mu_Y X \;+\; \mu_X\mu_Y] \\ &=\; E[XY] \;-\; \mu_X\mu_Y \;-\; \mu_Y\mu_X \;+\; \mu_X\mu_Y \\ &=\; E[XY] \;-\; E[X]\;E[Y]\;. \end{aligned}$$

$$
\begin{aligned}
Cov(X,Y) &\equiv E[\ (X-\mu_X)\ (Y-\mu_Y)\ ] \\
&= \sum_{k,\ell}\ (x_k-\mu_X)\ (y_\ell-\mu_Y)\ p(x_k,y_\ell) \\
&= E[XY]\ -\ E[X]\ E[Y]\ .
\end{aligned}
$$

**NOTE** :

$Cov(X,Y)$ measures "*concordance*" or "*coherence*" of $X$ and $Y$ :

- If $X > \mu_X$ when $Y > \mu_Y$ and $X < \mu_X$ when $Y < \mu_Y$ then

$$Cov(X,Y)\ >\ 0\ .$$

- If $X > \mu_X$ when $Y < \mu_Y$ and $X < \mu_X$ when $Y > \mu_Y$ then

$$Cov(X,Y)\ <\ 0\ .$$

**EXERCISE** : Prove the following :

- $Var(aX + b) \;=\; a^2\, Var(X)$ ,
- $Cov(X, Y) \;=\; Cov(Y, X)$ ,
- $Cov(cX, Y) \;=\; c\, Cov(X, Y)$ ,
- $Cov(X, cY) \;=\; c\, Cov(X, Y)$ ,
- $Cov(X + Y, Z) \;=\; Cov(X, Z) \;+\; Cov(Y, Z)$ ,
- $Var(X + Y) \;=\; Var(X) \;+\; Var(Y) \;+\; 2\, Cov(X, Y)$ .

**PROPERTY** :

If $X$ and $Y$ are *independent* then $Cov(X,Y) = 0$ .

**PROOF** :

We have already shown ( with $\mu_X \equiv E[X]$ and $\mu_Y \equiv E[Y]$ ) that

$$Cov(X,Y) \equiv E[\,(X-\mu_X)\,(Y-\mu_Y)\,] = E[XY] - E[X]\,E[Y]\,,$$

and that if $X$ and $Y$ are *independent* then

$$E[XY] = E[X]\,E[Y]\,.$$

from which the result follows.

**EXERCISE** : (already used earlier $\cdots$)

Probability mass function $p_{X,Y}(x,y)$

| | $y=6$ | $y=8$ | $y=10$ | $p_X(x)$ |
|---|---|---|---|---|
| $x=1$ | $\frac{1}{5}$ | $0$ | $\frac{1}{5}$ | $\frac{2}{5}$ |
| $x=2$ | $0$ | $\frac{1}{5}$ | $0$ | $\frac{1}{5}$ |
| $x=3$ | $\frac{1}{5}$ | $0$ | $\frac{1}{5}$ | $\frac{2}{5}$ |
| $p_Y(y)$ | $\frac{2}{5}$ | $\frac{1}{5}$ | $\frac{2}{5}$ | $1$ |

Show that

- $E[X] = 2 \quad , \quad E[Y] = 8 \quad , \quad E[XY] = 16$
- $Cov(X,Y) = E[XY] - E[X]\,E[Y] = 0$
- $X$ and $Y$ are *not* independent

Thus if

$$Cov(X,Y) = 0 \, ,$$

then it does not necessarily follow that $X$ and $Y$ are independent !

**PROPERTY** :

If $X$ and $Y$ are *independent* then

$$Var(X+Y) \;=\; Var(X) \;+\; Var(Y)\;.$$

**PROOF** :

We have already shown (in an exercise!) that

$$Var(X+Y) \;=\; Var(X) \;+\; Var(Y) \;+\; 2\,Cov(X,Y)\;,$$

and that if $X$ and $Y$ are *independent* then

$$Cov(X,Y) \;=\; 0\;,$$

from which the result follows.

**EXERCISE** :

Compute

$$E[X] \quad , \quad E[Y] \quad , \quad E[X^2] \quad , \quad E[Y^2]$$

$$E[XY] \quad , \quad Var(X) \quad , \quad Var(Y)$$

$$Cov(X, Y)$$

for

Joint probability mass function $p_{X,Y}(x, y)$

| | $y=0$ | $y=1$ | $y=2$ | $y=3$ | $p_X(x)$ |
|---|---|---|---|---|---|
| $x=0$ | $\frac{1}{8}$ | $0$ | $0$ | $0$ | $\frac{1}{8}$ |
| $x=1$ | $0$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{3}{8}$ |
| $x=2$ | $0$ | $\frac{2}{8}$ | $\frac{1}{8}$ | $0$ | $\frac{3}{8}$ |
| $x=3$ | $0$ | $\frac{1}{8}$ | $0$ | $0$ | $\frac{1}{8}$ |
| $p_Y(y)$ | $\frac{1}{8}$ | $\frac{4}{8}$ | $\frac{2}{8}$ | $\frac{1}{8}$ | $1$ |

**EXERCISE** :

Compute

$$E[X] \quad , \quad E[Y] \quad , \quad E[X^2] \quad , \quad E[Y^2]$$

$$E[XY] \quad , \quad Var(X) \quad , \quad Var(Y)$$

$$Cov(X, Y)$$

for

Joint probability mass function $p_{X,Y}(x, y)$

| | $y = 1$ | $y = 2$ | $y = 3$ | $p_X(x)$ |
|---|---|---|---|---|
| $x = 1$ | $\frac{1}{3}$ | $\frac{1}{12}$ | $\frac{1}{12}$ | $\frac{1}{2}$ |
| $x = 2$ | $\frac{2}{9}$ | $\frac{1}{18}$ | $\frac{1}{18}$ | $\frac{1}{3}$ |
| $x = 3$ | $\frac{1}{9}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{6}$ |
| $p_Y(y)$ | $\frac{2}{3}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | 1 |

# SPECIAL DISCRETE RANDOM VARIABLES

## The Bernoulli Random Variable

A *Bernoulli trial* has only *two outcomes*, with probability

$$P(X = 1) \quad = \quad p \ ,$$

$$P(X = 0) \quad = \quad 1 - p \ ,$$

*e.g.*, tossing a coin, winning or losing a game, $\cdots$ .

We have

$$E[X] \quad = \quad 1 \cdot p \ + \ 0 \cdot (1 - p) \quad = \quad p \ ,$$

$$E[X^2] \quad = \quad 1^2 \cdot p \ + \ 0^2 \cdot (1 - p) \quad = \quad p \ ,$$

$$Var(X) \quad = \quad E[X^2] - E[X]^2 \quad = \quad p - p^2 \quad = \quad p(1 - p) \ .$$

**NOTE** : If $p$ *is small* then $Var(X) \cong p$ .

**EXAMPLES** :

- When $p = \frac{1}{2}$ (*e.g.*, for tossing a coin), we have

$$E[X] \;=\; p \;=\; \tfrac{1}{2} \quad , \quad Var(X) \;=\; p(1-p) \;=\; \tfrac{1}{4} \; .$$

- When *rolling a die* , with outcome $k$ , $(1 \leq k \leq 6)$ , let

$$X(k) = 1 \;\; \text{if the roll resulted in a} \;\; \textit{six} \; ,$$

and

$$X(k) = 0 \;\; \text{if the roll did} \;\; \textit{not} \;\; \text{result in a} \;\; \textit{six} \; .$$

Then

$$E[X] \;=\; p \;=\; \tfrac{1}{6} \quad , \quad Var(X) \;=\; p(1-p) \;=\; \tfrac{5}{36} \; .$$

- When $p = 0.01$ , then

$$E[X] \;=\; 0.01 \quad , \quad Var(X) \;=\; 0.0099 \;\cong\; 0.01 \; .$$

## The Binomial Random Variable

Perform a Bernoulli trial $n$ times *in sequence* .

Assume the individual trials are *independent* .

An *outcome* could be

$$100011001010 \qquad (n = 12) \ ,$$

with probability

$$P(100011001010) \quad = \quad p^5 \cdot (1-p)^7 \ . \qquad (\text{ Why ? })$$

Let the $X$ be the number of "*successes* " (*i.e.* 1's) .

For example,

$$X(100011001010) \quad = \quad 5 \ .$$

We have

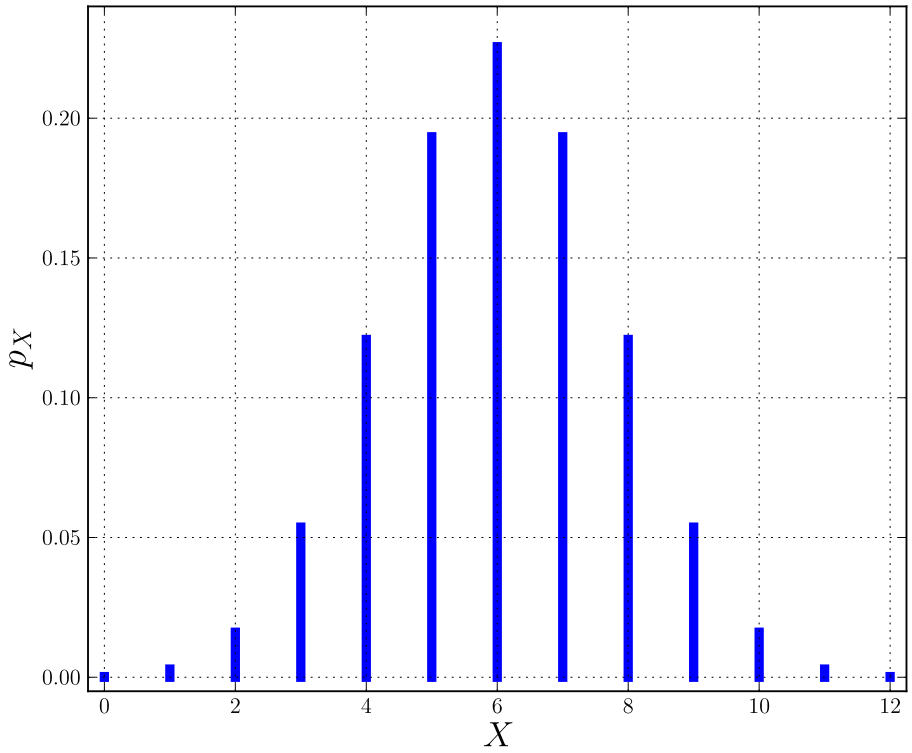
$$P(X = 5) \quad = \quad \binom{12}{5} \cdot p^5 \cdot (1-p)^7 \ . \qquad (\text{ Why ? })$$

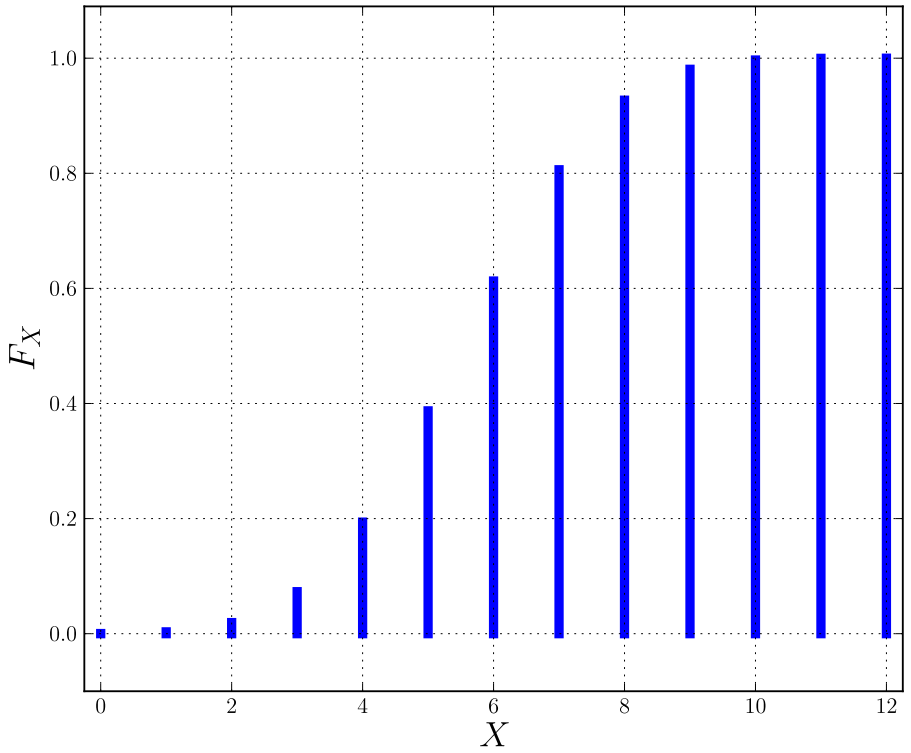
In general, for $k$ successes in a sequence of $n$ trials, we have

$$P(X = k) \quad = \quad \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \ , \qquad\qquad (0 \le k \le n) \ .$$

**EXAMPLE** : $\qquad n = 12 \quad , \quad \mathbf{p = \frac{1}{2}}$

| $k$ | $p_X(k)$ | $F_X(k)$ |
|---|---|---|
| 0 | 1 / 4096 | 1 / 4096 |
| 1 | 12 / 4096 | 13 / 4096 |
| 2 | 66 / 4096 | 79 / 4096 |
| 3 | 220 / 4096 | 299 / 4096 |
| 4 | 495 / 4096 | 794 / 4096 |
| 5 | 792 / 4096 | 1586 / 4096 |
| 6 | 924 / 4096 | 2510 / 4096 |
| 7 | 792 / 4096 | 3302 / 4096 |
| 8 | 495 / 4096 | 3797 / 4096 |
| 9 | 220 / 4096 | 4017 / 4096 |
| 10 | 66 / 4096 | 4083 / 4096 |
| 11 | 12 / 4096 | 4095 / 4096 |
| 12 | 1 / 4096 | 4096 / 4096 |



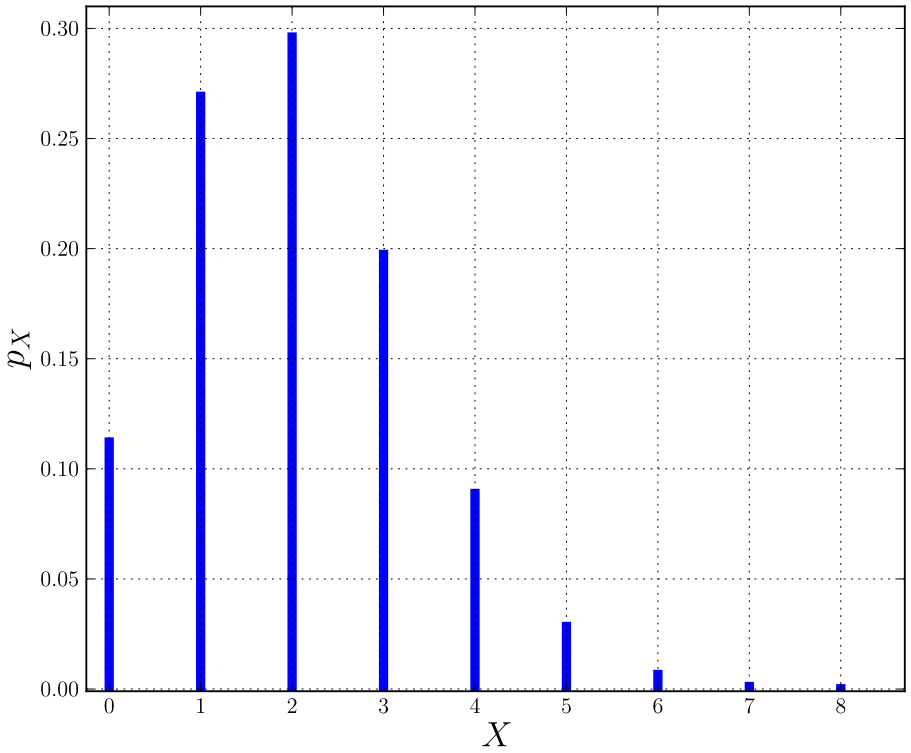

The Binomial *mass* and *distribution* functions for $n = 12$ , $p = \frac{1}{2}$

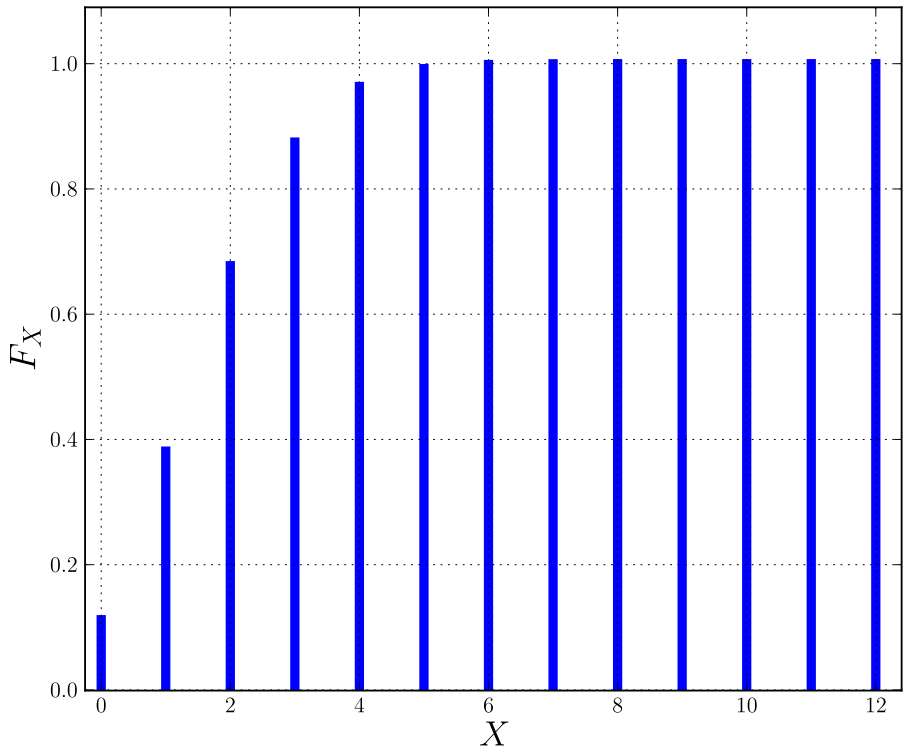
For $k$ successes in a sequence of $n$ trials :

$$P(X = k) \quad = \quad \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \ , \qquad\qquad (0 \leq k \leq n) \ .$$

**EXAMPLE** : $\qquad\qquad n = 12 \quad , \quad \mathbf{p = \frac{1}{6}}$

| $k$ | $p_X(k)$ | $F_X(k)$ |
|---|---|---|
| 0 | 0.1121566221 | 0.112156 |
| 1 | 0.2691758871 | 0.381332 |
| 2 | 0.2960935235 | 0.677426 |
| 3 | 0.1973956972 | 0.874821 |
| 4 | 0.0888280571 | 0.963649 |
| 5 | 0.0284249838 | 0.992074 |
| 6 | 0.0066324966 | 0.998707 |
| 7 | 0.0011369995 | 0.999844 |
| 8 | 0.0001421249 | 0.999986 |
| 9 | 0.0000126333 | 0.999998 |
| 10 | 0.0000007580 | 0.999999 |
| 11 | 0.0000000276 | 0.999999 |
| 12 | 0.0000000005 | 1.000000 |




The Binomial *mass* and *distribution* functions for $n = 12$ , $p = \frac{1}{6}$

**EXAMPLE** :

In 12 *rolls of a die* write the outcome as, for example,

$$100011001010$$

where

1 denotes the roll resulted in a *six* ,

and

0 denotes the roll did *not* result in a *six* .

As before, let $X$ be the number of 1's in the outcome.

Then $X$ represents the *number of sixes* in the 12 rolls.

Then, for example, using the preceding *Table* :

$$P(X = 5) \cong 2.8\ \% \quad , \quad P(X \leq 5) \cong 99.2\ \% \ .$$

**EXERCISE** : Show that from

$$P(X = k) \quad = \quad \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \ ,$$

and

$$P(X = k+1) \quad = \quad \binom{n}{k+1} \cdot p^{k+1} \cdot (1-p)^{n-k-1} \ ,$$

it follows that

$$P(X = k+1) \quad = \quad c_k \cdot P(X = k) \ ,$$

where

$$c_k \quad = \quad \frac{n-k}{k+1} \cdot \frac{p}{1-p} \quad .$$

**NOTE** : This *recurrence formula* is an efficient and stable *algorithm* to compute the binomial probabilities :

$$P(X = 0) \ = \ (1-p)^n \ ,$$

$$P(X = k+1) \quad = \quad c_k \cdot P(X = k) \ , \qquad k = 0, 1, \cdots, n-1 \ .$$

**Mean and variance of the Binomial random variable** :

By definition, the *mean* of a Binomial random variable $X$ is

$$E[X] \;=\; \sum_{k=0}^{n} k \cdot P(X=k) \;=\; \sum_{k=0}^{n} k \cdot \binom{n}{k} \; p^k \; (1-p)^{n-k} \;,$$

which can be shown to equal $np$ .

An *easy way* to see this is as follows :

If in a *sequence* of $n$ independent Bernoulli trials we let

$X_k \;=\;$ the outcome of the $k^{\text{th}}$ Bernoulli trial , $\quad (X_k = 0 \text{ or } 1)$ ,

then

$$X \;\equiv\; X_1 \,+\, X_2 \,+\, \cdots \,+\, X_n \;,$$

is the *Binomial random variable* that *counts the successes* " .

$$X \equiv X_1 + X_2 + \cdots + X_n$$

We know that

$$E[X_k] = p \, ,$$

so

$$E[X] = E[X_1] + E[X_2] + \cdots + E[X_n] = np \, .$$

We already know that

$$Var(X_k) = E[X_k^2] - (E[X_k])^2 = p - p^2 = p(1-p) \, ,$$

so, since the $X_k$ are *independent* , we have

$$Var(X) = Var(X_1) + Var(X_2) + \cdots + Var(X_n) = np(1-p) \, .$$

**NOTE** : If $p$ *is small* then $Var(X) \cong np$ .

**EXAMPLES** :

- For 12 tosses of a *coin* , with *Heads* is *success*, we have

so $$n = 12 \quad , \quad p = \frac{1}{2}$$

$$E[X] \;=\; np \;=\; 6 \quad , \quad Var(X) \;=\; np(1-p) \;=\; 3 \; .$$

- For 12 rolls of a *die* , with *six* is *success* , we have

so $$n = 12 \quad , \quad p = \frac{1}{6}$$

$$E[X] \;=\; np \;=\; 2 \quad , \quad Var(X) \;=\; np(1-p) \;=\; 5/3 \; .$$

- If $n = 500$ and $p = 0.01$ , then

$$E[X] \;=\; np \;=\; 5 \quad , \quad Var(X) \;=\; np(1-p) \;=\; 4.95 \;\cong\; 5 \; .$$

## The Poisson Random Variable

The Poisson variable *approximates* the Binomial random variable :

$$P(X = k) \quad = \quad \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad \cong \quad e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad ,$$

when we take

$$\lambda \quad = \quad n\, p \quad (\textit{the average number of successes}).$$

This approximation is *accurate* if $n$ is *large* and $p$ *small* .

Recall that for the Binomial random variable

$E[X] = n\, p$ , and $Var(X) = np(1-p) \cong np$ when $p$ is small.

Indeed, for the Poisson random variable we will show that

$$E[X] = \lambda \text{ and } Var(X) = \lambda \ .$$

A *stable* and *efficient* way to compute the Poisson probability

$$P(X = k) \quad = \quad e^{-\lambda} \cdot \frac{\lambda^k}{k!} , \qquad k = 0, 1, 2, \ \cdots \ ,$$

$$P(X = k+1) \quad = \quad e^{-\lambda} \cdot \frac{\lambda^{k+1}}{(k+1)!} ,$$

is to use the *recurrence relation*

$$P(X = 0) \ = \ e^{-\lambda} ,$$

$$P(X = k+1) \quad = \quad \frac{\lambda}{k+1} \cdot P(X = k) , \qquad k = 0, 1, 2, \ \cdots \ .$$

**NOTE** : Unlike the Binomial random variable, the Poisson random variable can have any nonnegative integer value $k$.

The Poisson random variable

$$P(X = k) \quad = \quad e^{-\lambda} \cdot \frac{\lambda^k}{k!} \, , \qquad k = 0, 1, 2, \cdots \, ,$$

has (as shown later) : $\quad E[X] \; = \; \lambda \quad$ and $\quad Var(X) \; = \; \lambda \, .$

The Poisson *distribution function* is

$$F(k) \quad = \quad P(X \leq k) \quad = \quad \sum_{\ell=0}^{k} e^{-\lambda} \, \frac{\lambda^\ell}{\ell!} \quad = \quad e^{-\lambda} \sum_{\ell=0}^{k} \frac{\lambda^\ell}{\ell!} \, ,$$

with, as should be the case,

$$\lim_{k \to \infty} \; F(k) \quad = \quad e^{-\lambda} \sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{\ell!} \quad = \quad e^{-\lambda} \, e^{\lambda} \quad = \quad 1 \, .$$

(using the *Taylor series* from Calculus for $e^\lambda$) .

The Poisson random variable

$$P(X = k) \quad = \quad e^{-\lambda} \cdot \frac{\lambda^k}{k!} , \qquad k = 0, 1, 2, \ \cdots \ ,$$

models the probability of $k$ "*successes*" in a given "time" interval, when the *average* number of successes is $\lambda$ .

**EXAMPLE** : Suppose customers arrive at the rate of *six* per hour. The probability that $k$ customers arrive in a one-hour period is

$$P(k = 0) \quad = \quad e^{-6} \cdot \frac{6^0}{0!} \quad \cong \quad 0.0024 ,$$

$$P(k = 1) \quad = \quad e^{-6} \cdot \frac{6^1}{1!} \quad \cong \quad 0.0148 ,$$

$$P(k = 2) \quad = \quad e^{-6} \cdot \frac{6^2}{2!} \quad \cong \quad 0.0446 .$$

The probability that more than 2 customers arrive is

$$1 \ - \ (0.0024 \ + \ 0.0148 \ + \ 0.0446) \quad \cong \quad 0.938 .$$

$$p_{\text{Binomial}} = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad \cong \quad p_{\text{Poisson}} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

**EXAMPLE** : $\lambda = 6$ customers/hour.

For the Binomial take $n = 12$ , $p = 0.5$ (0.5 customers/5 minutes) ,

so that indeed $np = \lambda$ .

| $k$ | $p_{\text{Binomial}}$ | $p_{\text{Poisson}}$ | $F_{\text{Binomial}}$ | $F_{\text{Poisson}}$ |
|---|---|---|---|---|
| 0 | 0.0002 | 0.0024 | 0.0002 | 0.0024 |
| 1 | 0.0029 | 0.0148 | 0.0031 | 0.0173 |
| 2 | 0.0161 | 0.0446 | 0.0192 | 0.0619 |
| 3 | 0.0537 | 0.0892 | 0.0729 | 0.1512 |
| 4 | 0.1208 | 0.1338 | 0.1938 | 0.2850 |
| 5 | 0.1933 | 0.1606 | 0.3872 | 0.4456 |
| 6 | 0.2255 | 0.1606 | 0.6127 | 0.6063 |
| 7 | 0.1933 | 0.1376 | 0.8061 | 0.7439 |
| 8 | 0.1208 | 0.1032 | 0.9270 | 0.8472 |
| 9 | 0.0537 | 0.0688 | 0.9807 | 0.9160 |
| 10 | 0.0161 | 0.0413 | 0.9968 | 0.9573 |
| 11 | 0.0029 | 0.0225 | 0.9997 | 0.9799 |
| 12 | 0.0002 | 0.0112 | 1.0000 | 0.9911$\star$ |

Here the approximation is *not so good* $\cdots$

$$p_{\text{Binomial}} = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad \cong \quad p_{\text{Poisson}} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

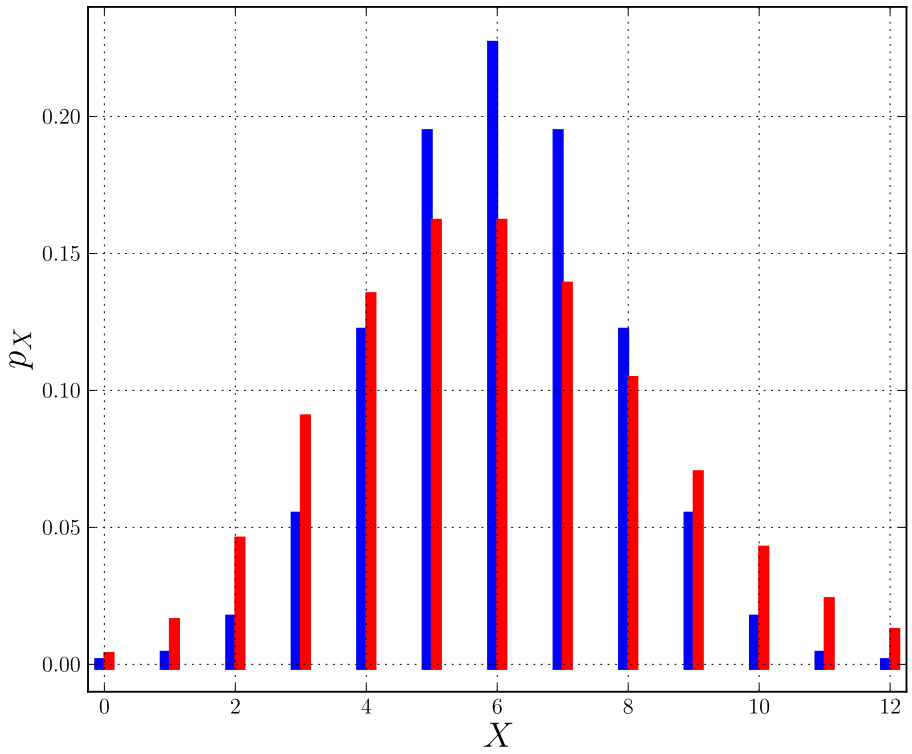
EXAMPLE : $\lambda = 6$ customers/hour.

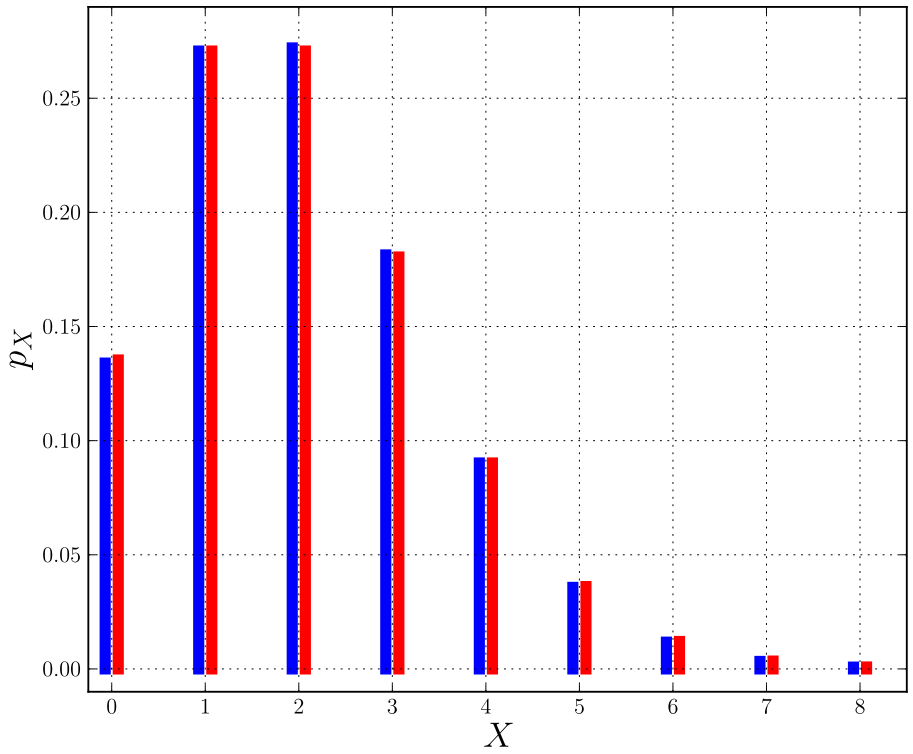
For the Binomial take $n = 60$ , $p = 0.1$ (0.1 customers/minute) ,

so that indeed $np = \lambda$ .

| $k$ | $p_{\text{Binomial}}$ | $p_{\text{Poisson}}$ | $F_{\text{Binomial}}$ | $F_{\text{Poisson}}$ |
|---|---|---|---|---|
| 0 | 0.0017 | 0.0024 | 0.0017 | 0.0024 |
| 1 | 0.0119 | 0.0148 | 0.0137 | 0.0173 |
| 2 | 0.0392 | 0.0446 | 0.0530 | 0.0619 |
| 3 | 0.0843 | 0.0892 | 0.1373 | 0.1512 |
| 4 | 0.1335 | 0.1338 | 0.2709 | 0.2850 |
| 5 | 0.1662 | 0.1606 | 0.4371 | 0.4456 |
| 6 | 0.1692 | 0.1606 | 0.6064 | 0.6063 |
| 7 | 0.1451 | 0.1376 | 0.7515 | 0.7439 |
| 8 | 0.1068 | 0.1032 | 0.8583 | 0.8472 |
| 9 | 0.0685 | 0.0688 | 0.9269 | 0.9160 |
| 10 | 0.0388 | 0.0413 | 0.9657 | 0.9573 |
| 11 | 0.0196 | 0.0225 | 0.9854 | 0.9799 |
| 12 | 0.0089 | 0.0112 | 0.9943 | 0.9911 |
| 13 | ... | ... | ... | ... |

Here the approximation is *better* ...




$n = 12 \quad , \quad p = \frac{1}{2} \quad , \quad \lambda = 6$

$n = 200 \quad , \quad p = 0.01 \; , \quad \lambda = 2$

The Binomial (*blue*) and Poisson (*red*) probability mass functions.

For the case $n = 200, \; p = 0.01$, the approximation is very good !

For the *Binomial* random variable we found

$$E[X] \;=\; np \qquad \text{and} \qquad Var(X) \;=\; np(1-p)\;,$$

while for the *Poisson* random variable, with $\lambda = np$ we will show

$$E[X] \;=\; np \qquad \text{and} \qquad Var(X) \;=\; np\;.$$

Note again that

$$np(1-p) \;\cong\; np \quad , \qquad \text{when } p \text{ is } \textit{small}\;.$$

**EXAMPLE** : In the preceding two *Tables* we have

n=12 , p=0.5

| | Binomial | Poisson |
|---|---|---|
| $E[X]$ | 6.0000 | 6.0000 |
| $Var[X]$ | 3.0000 | 6.0000 |
| $\sigma[X]$ | 1.7321 | 2.4495 |

n=60 , p=0.1

| | Binomial | Poisson |
|---|---|---|
| $E[X]$ | 6.0000 | 6.0000 |
| $Var[X]$ | 5.4000 | 6.0000 |
| $\sigma[X]$ | 2.3238 | 2.4495 |

**FACT** : (*The Method of Moments*)

By *Taylor expansion* of $e^{tX}$ about $t = 0$ , we have

$$\psi(t) \equiv E[e^{tX}] = E[\,1 + tX + \frac{t^2\,X^2}{2!} + \frac{t^3\,X^3}{3!} + \cdots\,]$$

$$= 1 + t\,E[X] + \frac{t^2}{2!}\,E[X^2] + \frac{t^3}{3!}\,E[X^3] + \cdots .$$

It follows that

$$\psi'(0) = E[X] \quad , \quad \psi''(0) = E[X^2] \quad , \quad etc. ,$$

which sometimes *facilitates computing the mean*

$$\mu = E[X] ,$$

*and the variance*

$$Var(X) = E[X^2] - \mu^2 .$$

**APPLICATION** : The *Poisson mean* and *variance* :

$$\psi(t) \equiv E[e^{tX}] = \sum_{k=0}^{\infty} e^{tk}\ P(X=k) = \sum_{k=0}^{\infty} e^{tk}\ e^{-\lambda}\ \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda}\ e^{\lambda\ e^t} = e^{\lambda(e^t-1)}\ .$$

Here

$$\psi'(t) = \lambda\ e^t\ e^{\lambda(e^t-1)}$$

$$\psi''(t) = \lambda\ \left[\lambda\ (e^t)^2\ +\ e^t\right]\ e^{\lambda(e^t-1)}$$

so that

$$E[X] = \psi'(0) = \lambda$$

$$E[X^2] = \psi''(0) = \lambda(\lambda+1) = \lambda^2\ +\ \lambda$$

$$Var(X) = E[X^2]\ -\ E[X]^2 = \lambda\ .$$

**EXAMPLE** : *Defects* in a wire occur at the rate of *one per* 10 *meter*, with a *Poisson distribution* :

$$P(X = k) \quad = \quad e^{-\lambda} \cdot \frac{\lambda^k}{k!} \, , \qquad k = 0, 1, 2, \ \cdots \ .$$

What is the probability that :

- A 12-meter roll has at *no* defects?

  **ANSWER** : Here $\lambda = 1.2$ , and $P(X = 0) \ = \ e^{-\lambda} \ = \ 0.3012$ .

- A 12-meter roll of wire has *one* defect?

  **ANSWER** : With $\lambda = 1.2$ , $P(X = 1) \ = \ e^{-\lambda} \cdot \lambda \ = \ 0.3614$ .

- Of *five* 12-meter rolls *two* have *one* defect and *three* have *none*?

  **ANSWER** : $\binom{5}{3} \cdot 0.3012^3 \cdot 0.3614^2 \ = \ 0.0357$ . ( Why ? )

**EXERCISE** :

Defects in a certain wire occur at the rate of one per 10 meter.
Assume the defects have a Poisson distribution.
What is the probability that :

- a 20-meter wire has no defects?
- a 20-meter wire has at most 2 defects?

**EXERCISE** :

Customers arrive at a counter at the rate of 8 per hour.
Assume the arrivals have a Poisson distribution.
What is the probability that :

- no customer arrives in 15 minutes?
- two customers arrive in a period of 30 minutes?

# CONTINUOUS RANDOM VARIABLES

**DEFINITION** : A *continuous random variable* is a *function* $X(s)$ from an *uncountably infinite* sample space $\mathcal{S}$ to the real numbers $\mathbb{R}$ ,

$$X(\cdot) \quad : \quad \mathcal{S} \quad \rightarrow \quad \mathbb{R} \, .$$

**EXAMPLE** :

Rotate a *pointer* about a pivot in a plane (like a hand of a clock).

The *outcome* is the *angle* where it stops : $2\pi\theta$ , where $\theta \in (0, 1]$ .

A good *sample space* is all values of $\theta$ , *i.e.* $\mathcal{S} = (0, 1]$ .

A very simple example of a *continuous random variable* is $X(\theta) = \theta$ .

Suppose *any outcome*, *i.e.*, any value of $\theta$ is "equally likely".

What are the values of

$$P(0 < \theta \leq \tfrac{1}{2}) \quad , \quad P(\tfrac{1}{3} < \theta \leq \tfrac{1}{2}) \quad , \quad P(\theta = \tfrac{1}{\sqrt{2}}) \; ?$$

The (*cumulative*) *probability distribution function* is defined as

$$F_X(x) \equiv P(X \leq x) .$$

Thus

$$F_X(b) - F_X(a) \equiv P(a < X \leq b) .$$

We must have

$$F_X(-\infty) = 0 \quad \text{and} \quad F_X(\infty) = 1 ,$$

*i.e.*,

$$\lim_{x \to -\infty} F_X(x) = 0 ,$$

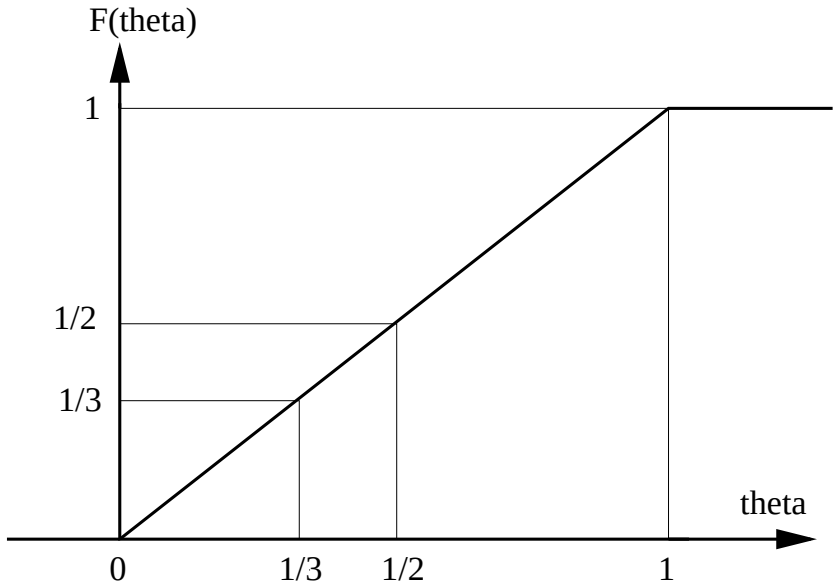
and

$$\lim_{x \to \infty} F_X(x) = 1 .$$

Also, $F_X(x)$ is a *non-decreasing* function of $x$ . ( Why ? )

**NOTE** : All the above is *the same* as for *discrete* random variables !

**EXAMPLE** : In the "*pointer example*", where $X(\theta) = \theta$ , we have the *probability distribution function*




Note that

$$F(\tfrac{1}{3}) \equiv P(X \leq \tfrac{1}{3}) = \tfrac{1}{3} \quad , \quad F(\tfrac{1}{2}) \equiv P(X \leq \tfrac{1}{2}) = \tfrac{1}{2} \quad ,$$

$$P(\tfrac{1}{3} < X \leq \tfrac{1}{2}) = F(\tfrac{1}{2}) - F(\tfrac{1}{3}) = \tfrac{1}{2} - \tfrac{1}{3} = \tfrac{1}{6} \, .$$

**QUESTION** : What is $P(\tfrac{1}{3} \leq X \leq \tfrac{1}{2})$ ?

The *probability density function* is the *derivative* of the probability distribution function :

$$f_X(x) \equiv F'_X(x) \equiv \frac{d}{dx} F_X(x) .$$

**EXAMPLE** : In the "*pointer example* "

$$F_X(x) = \begin{cases} 0 , & x \leq 0 \\ x , & 0 < x \leq 1 \\ 1 , & 1 < x \end{cases}$$

Thus

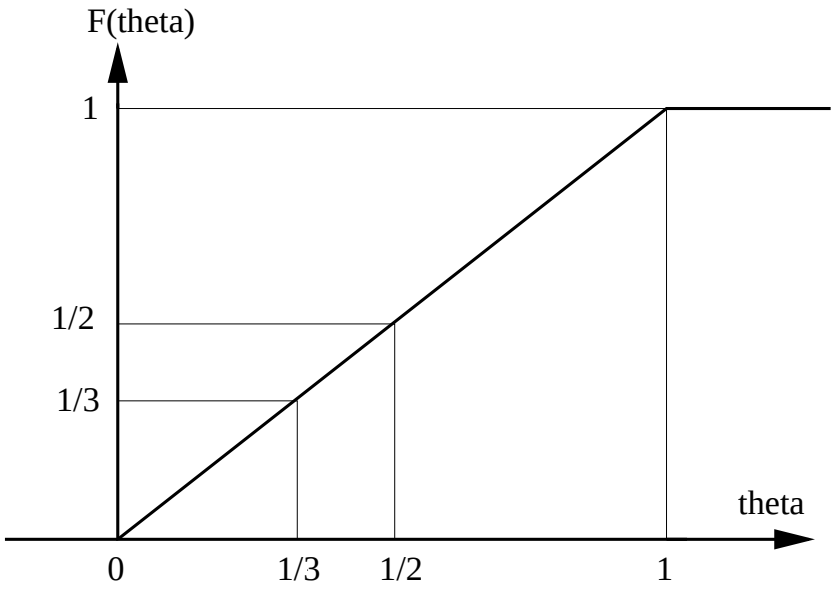
$$f_X(x) = F'_X(x) = \begin{cases} 0 , & x \leq 0 \\ 1 , & 0 < x \leq 1 \\ 0 , & 1 < x \end{cases}$$

**NOTATION** : When it is clear what $X$ is then we also write

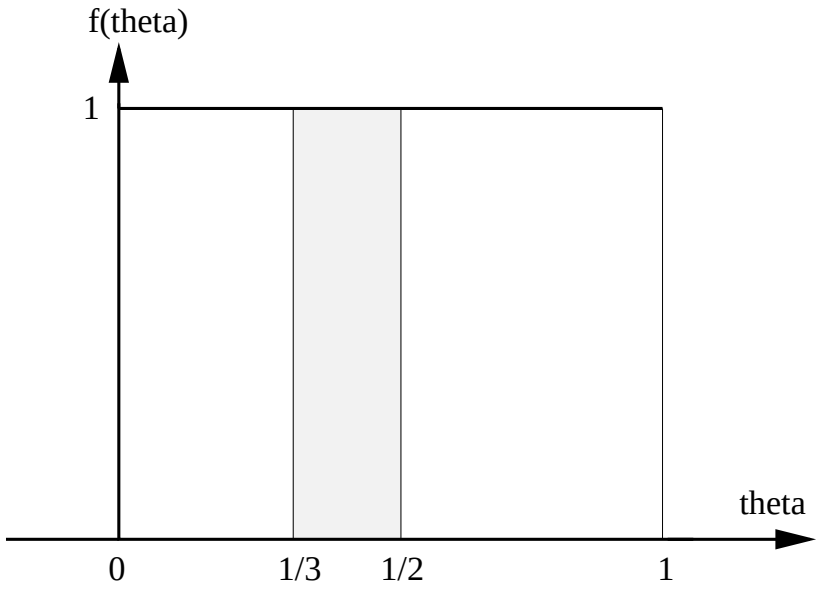
$$f(x) \quad \text{for} \quad f_X(x) , \qquad \text{and} \qquad F(x) \quad \text{for} \quad F_X(x) .$$

**EXAMPLE** : ( continued $\cdots$ )

$$F(x) = \begin{cases} 0\ , & x \le 0 \\ x\ , & 0 < x \le 1 \\ 1\ , & 1 < x \end{cases} \quad , \quad f(x) = \begin{cases} 0\ , & x \le 0 \\ 1\ , & 0 < x \le 1 \\ 0\ , & 1 < x \end{cases}$$




Distribution function




Density function

**NOTE** :

$$P(\frac{1}{3} < X \le \frac{1}{2}) \quad = \quad \int_{\frac{1}{3}}^{\frac{1}{2}} f(x)\ dx \quad = \quad \frac{1}{6} \quad = \quad \text{the shaded area .}$$

In general, from

$$f(x) \quad \equiv \quad F'(x) \ ,$$

with

$$F(-\infty) \quad = \quad 0 \qquad \text{and} \qquad F(\infty) \quad = \quad 1 \ ,$$

we have from Calculus the following *basic identities* :

$$\int_{-\infty}^{\infty} f(x)\ dx \quad = \quad \int_{-\infty}^{\infty} F'(x)\ dx \quad = \quad F(\infty) - F(-\infty) \quad = \quad 1 \ ,$$

$$\int_{-\infty}^{x} f(x)\ dx \quad = \quad F(x) - F(-\infty) \ = \ F(x) \ = \ P(X \leq x) \ ,$$

$$\int_{a}^{b} f(x)\ dx \quad = \quad F(b) - F(a) \quad = \quad P(a < X \leq b) \ ,$$

$$\int_{a}^{a} f(x)\ dx \quad = \quad F(a) - F(a) \quad = \quad 0 \quad = \quad P(X = a) \ .$$

**EXERCISE** : Draw *graphs* of the distribution and density functions

$$F(x) = \begin{cases} 0\,, & x \le 0 \\ 1 - e^{-x}\,, & x > 0 \end{cases} \quad , \quad f(x) = \begin{cases} 0\,, & x \le 0 \\ e^{-x}\,, & x > 0 \end{cases} \quad ,$$

and verify that

- $F(-\infty) \;=\; 0\,, \qquad F(\infty) \;=\; 1\,,$
- $f(x) \;=\; F'(x)\,,$
- $F(x) \;=\; \int_0^x \, f(x)\, dx\,,$ ( Why is *zero* as lower limit OK ? )
- $\int_0^\infty f(x)\, dx \;=\; 1\,,$
- $P(0 < X \le 1) \;=\; F(1) - F(0) \;=\; F(1) \;=\; 1 - e^{-1} \;\cong\; 0.63\,,$
- $P(X > 1) \;=\; 1 - F(1) \;=\; e^{-1} \;\cong\; 0.37\,,$
- $P(1 < X \le 2) \;=\; F(2) - F(1) \;=\; e^{-1} - e^{-2} \;\cong\; 0.23\,.$

**EXERCISE** : For positive integer $n$, consider the density functions

$$f_n(x) = \begin{cases} cx^n(1-x^n)\ , & 0 \le x \le 1 \\ \\ 0\ , & \text{otherwise} \end{cases}$$

- Determine the value of $c$ in terms of $n$ .
- Draw the graph of $f_n(x)$ for $n = 1, 2, 4, 8, 16$ .
- Determine the distribution function $F_n(x)$ .
- Draw the graph of $F_n(x)$ for $n = 1, 2, 3, 4, 8, 16$ .
- Determine $P(0 \le X \le \frac{1}{2})$ in terms of $n$ .
- What happens to $P(0 \le X \le \frac{1}{2})$ when $n$ becomes large?
- Determine $P(\frac{9}{10} \le X \le 1)$ in terms of $n$ .
- What happens to $P(\frac{9}{10} \le X \le 1)$ when $n$ becomes large?

## Joint distributions

A *joint probability density function* $f_{X,Y}(x,y)$ must satisfy

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy \;=\; 1 \qquad (\text{ "Volume" } = 1\ ).$$

The corresponding *joint probability distribution function* is

$$F_{X,Y}(x,y) \;=\; P(X \le x\ ,\ Y \le y) \;=\; \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(x,y)\,dx\,dy\ .$$

By Calculus we have $$\frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \;=\; f_{X,Y}(x,y)\ .$$

Also,

$$P(a < X \le b\ ,\ c < Y \le d) \;=\; \int_{c}^{d}\int_{a}^{b} f_{X,Y}(x,y)\,dx\,dy\ .$$

**EXAMPLE** :

If

$$f_{X,Y}(x,y) \;=\; \begin{cases} 1 & \text{for } x \in (0,1] \text{ and } y \in (0,1] \ , \\ \\ 0 & \text{otherwise} \ , \end{cases}$$

then, for $x \in (0,1]$ and $y \in (0,1]$ ,

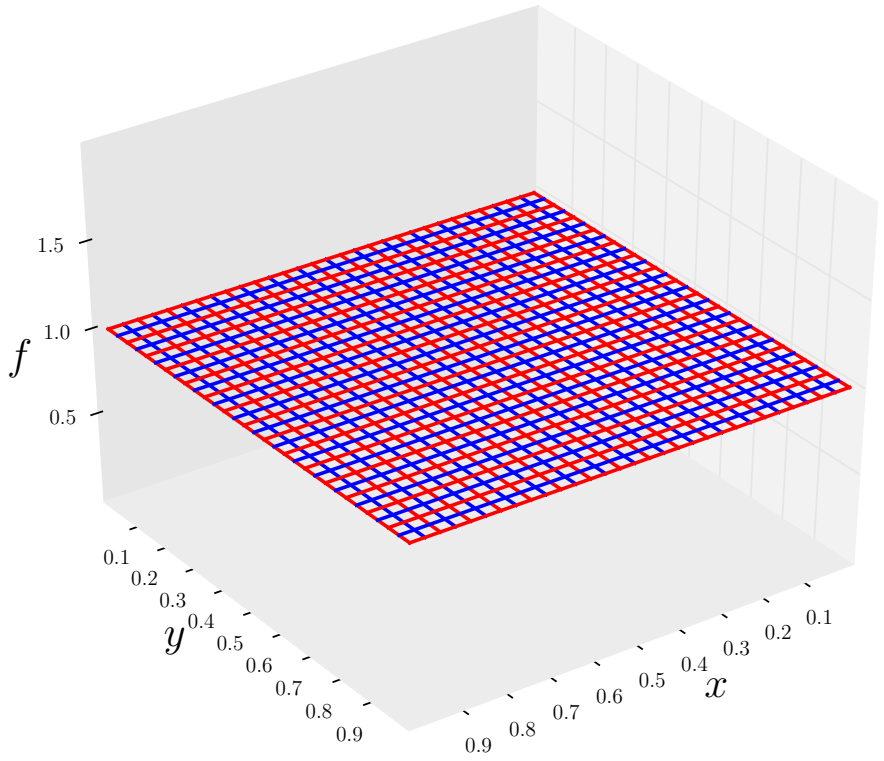
$$F_{X,Y}(x,y) \;=\; P(X \le x \ , \ Y \le y) \;=\; \int_0^y \int_0^x 1 \; dx \; dy \;=\; xy \ .$$

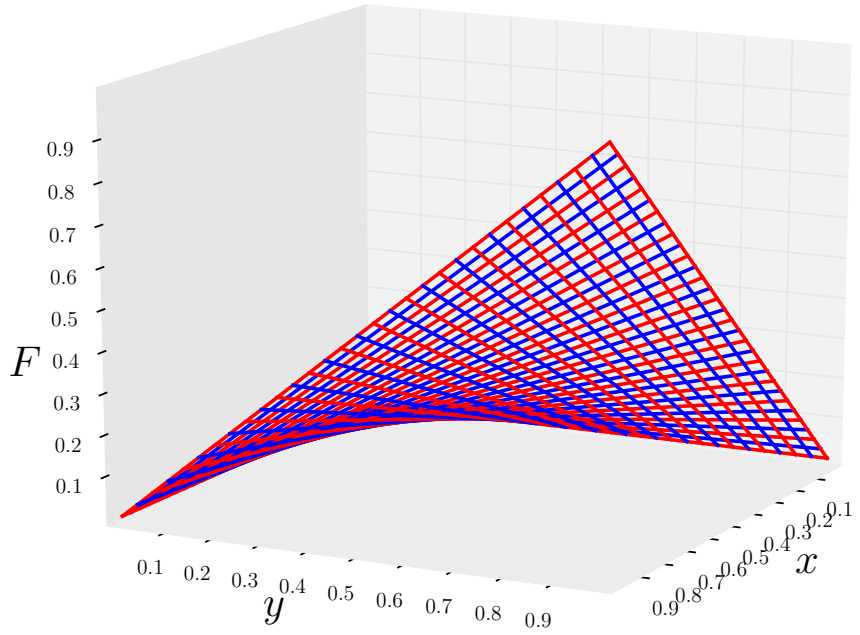
Thus

$$F_{X,Y}(x,y) \;=\; xy \ , \qquad \text{for } x \in (0,1] \text{ and } y \in (0,1] \ .$$

For example

$$P(X \le \frac{1}{3} \ , \ Y \le \frac{1}{2}) \;=\; F_{X,Y}(\frac{1}{3} \ , \ \frac{1}{2}) \;=\; \frac{1}{6} \ .$$




Also,

$$P(\,\frac{1}{3} \le X \le \frac{1}{2}\,,\,\frac{1}{4} \le Y \le \frac{3}{4}\,) \;=\; \int_{\frac{1}{4}}^{\frac{3}{4}} \int_{\frac{1}{3}}^{\frac{1}{2}} f(x,y)\,dx\,dy \;=\; \frac{1}{12}\,.$$

**EXERCISE** : Show that we can also compute this as follows :

$$F(\tfrac{1}{2},\tfrac{3}{4}) \;-\; F(\tfrac{1}{3},\tfrac{3}{4}) \;-\; F(\tfrac{1}{2},\tfrac{1}{4}) \;+\; F(\tfrac{1}{3},\tfrac{1}{4}) \;=\; \tfrac{1}{12}\,.$$

and explain why !

## Marginal density functions

The *marginal density functions* are

$$f_X(x) \;=\; \int_{-\infty}^{\infty} f_{X,Y}(x,y)\; dy \quad , \quad f_Y(y) \;=\; \int_{-\infty}^{\infty} f_{X,Y}(x,y)\; dx \;.$$

with corresponding *marginal distribution functions*

$$F_X(x) \;\equiv\; P(X \le x) \;=\; \int_{-\infty}^{x} f_X(x)\, dx \;=\; \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy\, dx \;,$$

$$F_Y(y) \;\equiv\; P(Y \le y) \;=\; \int_{-\infty}^{y} f_Y(y)\, dy \;=\; \int_{-\infty}^{y}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy \;.$$

By Calculus we have

$$\frac{dF_X(x)}{dx} \;=\; f_X(x) \qquad , \qquad \frac{dF_Y(y)}{dy} \;=\; f_Y(y) \;.$$

**EXAMPLE** : If

$$f_{X,Y}(x,y) \;=\; \begin{cases} 1 & \text{for } x \in (0,1] \text{ and } y \in (0,1] \;, \\ \\ 0 & \text{otherwise} \quad , \end{cases}$$

then, for $x \in (0,1]$ and $y \in (0,1]$ ,

$$f_X(x) \;=\; \int_0^1 f_{X,Y}(x,y)\; dy \;=\; \int_0^1 1\; dy \;=\; 1 \;,$$

$$f_Y(y) \;=\; \int_0^1 f_{X,Y}(x,y)\; dx \;=\; \int_0^1 1\; dx \;=\; 1 \;,$$

$$F_X(x) \;=\; P(X \le x) \;=\; \int_0^x f_X(x)\; dx \;=\; x \;,$$

$$F_Y(y) \;=\; P(Y \le y) \;=\; \int_0^y f_Y(y)\; dy \;=\; y \;.$$

For example

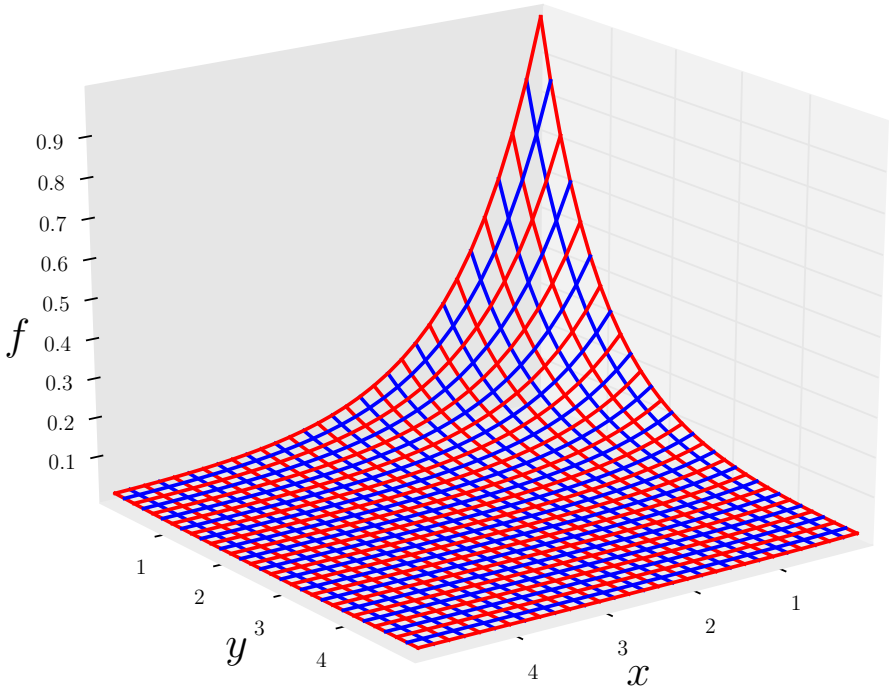
$$P(X \le \tfrac{1}{3}) \;=\; F_X(\tfrac{1}{3}) \;=\; \tfrac{1}{3} \quad , \quad P(Y \le \tfrac{1}{2}) \;=\; F_Y(\tfrac{1}{2}) \;=\; \tfrac{1}{2} \;.$$

**EXERCISE** :

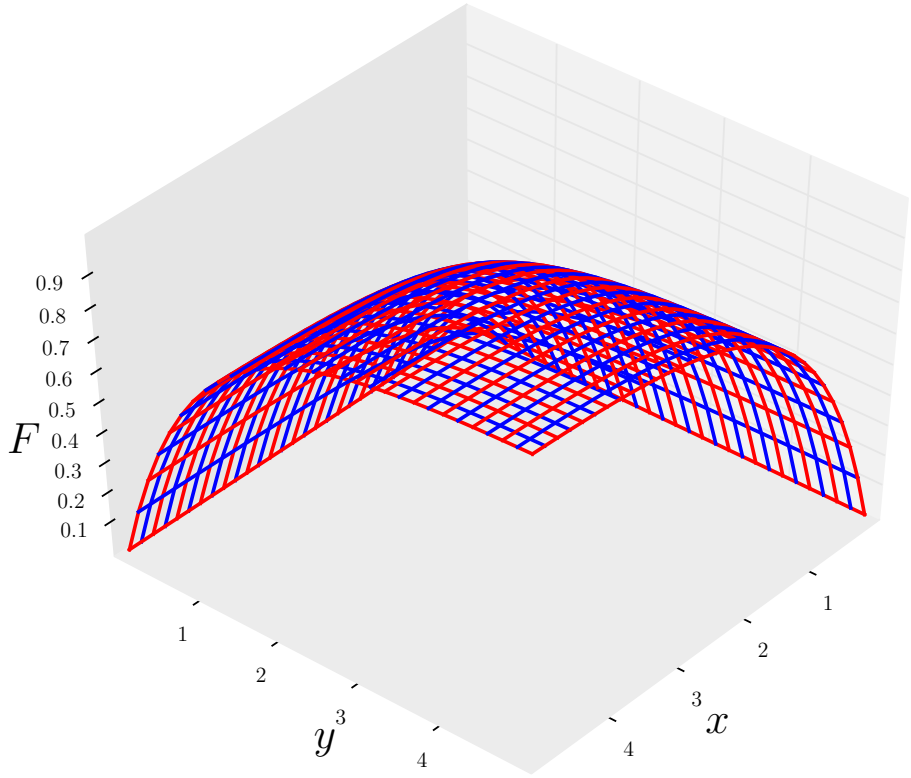
Let $F_{X,Y}(x,y) = \begin{cases} (1-e^{-x})(1-e^{-y}) & \text{for } x \geq 0 \text{ and } y \geq 0 \,, \\ 0 & \text{otherwise} \,. \end{cases}$

- Verify that

$$f_{X,Y}(x,y) = \frac{\partial^2 F}{\partial x \partial y} = \begin{cases} e^{-x-y} & \text{for } x \geq 0 \text{ and } y \geq 0 \,, \\ 0 & \text{otherwise} \,. \end{cases}$$




Density function $f_{X,Y}(x,y)$




Distribution function $F_{X,Y}(x,y)$

**EXERCISE** : ( continued $\cdots$ )

$F_{X,Y}(x,y) \;=\; (1-e^{-x})(1-e^{-y}) \quad , \quad f_{X,Y}(x,y) \;=\; e^{-x-y}\,, \quad \text{for } x,y \geq 0\,.$

Also verify the following :

- $F(0,0) = 0 \quad , \quad F(\infty,\infty) = 1\,,$
- $\int_0^\infty \int_0^\infty \; f_{X,Y}(x,y)\,dx\,dy \;=\; 1\,, \qquad$ ( Why *zero* lower limits ? )
- $f_X(x) \;=\; \int_0^\infty e^{-x-y}\,dy \;=\; e^{-x}\,,$
- $f_Y(y) \;=\; \int_0^\infty e^{-x-y}\,dx \;=\; e^{-y}\,.$
- $f_{X,Y}(x,y) \;=\; f_X(x) \;\cdot\; f_Y(y)\,. \qquad$ ( So ? )

**EXERCISE** : ( continued $\cdots$ )

$$F_{X,Y}(x,y) \;=\; (1-e^{-x})(1-e^{-y}) \quad , \quad f_{X,Y}(x,y) \;=\; e^{-x-y} \, , \quad \text{for } x,y \geq 0 \, .$$

Also verify the following :

- $F_X(x) \;=\; \int_0^x f_X(x)\, dx \;=\; \int_0^x e^{-x}\, dx \;=\; 1 - e^{-x}$ ,
- $F_Y(y) \;=\; \int_0^y f_Y(y)\, dy \;=\; \int_0^y e^{-y}\, dy \;=\; 1 - e^{-y}$ ,
- $F_{X,Y}(x,y) \;=\; F_X(x) \cdot F_Y(y)$ . ( So ? )
- $P(1 < x < \infty) \;=\; F_X(\infty) - F_X(1) \;=\; 1-(1-e^{-1}) \;=\; e^{-1} \;\cong\; 0.37$ ,
- $P(\; 1 < x \leq 2 \, , \; 0 < y \leq 1 \;) \;=\; \int_0^1 \int_1^2 \; e^{-x-y}\, dx\, dy$

$$= \; (e^{-1} - e^{-2})(1 - e^{-1}) \;\cong\; 0.15 \, ,$$

## Independent continuous random variables

Recall that two events $E$ and $F$ are *independent* if

$$P(EF) \quad = \quad P(E)\ P(F)\ .$$

Continuous random variables $X(s)$ and $Y(s)$ are *independent* if

$$P(X \in I_X\ ,\ Y \in I_Y) \quad = \quad P(X \in I_X) \cdot P(Y \in I_Y)\ ,$$

for *all* allowable sets $I_X$ and $I_Y$ (typically *intervals*) of *real numbers*.

Equivalently, $X(s)$ and $Y(s)$ are independent if for all such sets $I_X$ and $I_Y$ the *events*

$$X^{-1}(I_X) \quad \text{and} \quad Y^{-1}(I_Y)\ ,$$

are independent *in the sample space* $\mathcal{S}$.

**NOTE** :

$$X^{-1}(I_X) \equiv \{s \in \mathcal{S}\ :\ X(s) \in I_X\}\ ,$$
$$Y^{-1}(I_Y) \equiv \{s \in \mathcal{S}\ :\ Y(s) \in I_Y\}\ .$$

**FACT** : $X(s)$ and $Y(s)$ are *independent* if for all $x$ and $y$

$$f_{X,Y}(x,y) \quad = \quad f_X(x) \cdot f_Y(y) \; .$$

**EXAMPLE** : The random variables with density function

$$f_{X,Y}(x,y) \; = \; \begin{cases} e^{-x-y} & \text{for } x \geq 0 \text{ and } y \geq 0 \; , \\ 0 & \text{otherwise} \; , \end{cases}$$

are *independent* because (by the preceding exercise)

$$f_{X,Y}(x,y) \quad = \quad e^{-x-y} \quad = \quad e^{-x} \cdot e^{-y} \quad = \quad f_X(x) \cdot f_Y(y) \; .$$

**NOTE** :

$$F_{X,Y}(x,y) \; = \; \begin{cases} (1-e^{-x})(1-e^{-y}) & \text{for } x \geq 0 \text{ and } y \geq 0 \; , \\ 0 & \text{otherwise} \; , \end{cases}$$

also satisfies (by the preceding exercise)

$$F_{X,Y}(x,y) \quad = \quad F_X(x) \cdot F_Y(y) \; .$$

**PROPERTY** :

For *independent* continuous random variables $X$ and $Y$ we have

$$F_{X,Y}(x,y) \quad = \quad F_X(x) \cdot F_Y(y) \ , \qquad \text{for all } x, y \ .$$

**PROOF** :

$$\begin{aligned}
F_{X,Y}(x,y) &= P(X \leq x \ , \ Y \leq y) \\
&= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x,y) \ dy \ dx \\
&= \int_{-\infty}^{x} \int_{-\infty}^{y} f_X(x) \cdot f_Y(y) \ dy \ dx \qquad \text{(by independence)} \\
&= \int_{-\infty}^{x} [ \ f_X(x) \cdot \int_{-\infty}^{y} f_Y(y) \ dy \ ] \ dx \\
&= [ \ \int_{-\infty}^{x} f_X(x) \ dx \ ] \ \cdot \ [ \ \int_{-\infty}^{y} f_Y(y) \ dy \ ] \\
&= F_X(x) \cdot F_Y(y) \ .
\end{aligned}$$

**REMARK** : Note how the proof parallels that for the discrete case !

## Conditional distributions

Let $X$ and $Y$ be continuous random variables.

For given allowable sets $I_X$ and $I_Y$ (typically *intervals*), let

$$E_x \;=\; X^{-1}(I_X) \qquad \text{and} \qquad E_y \;=\; Y^{-1}(I_Y)\ ,$$

be their corresponding *events* in the sample space $\mathcal{S}$ .

We have $$P(E_x|E_y) \;\equiv\; \frac{P(E_xE_y)}{P(E_y)}\ .$$

The *conditional probability density function* is defined as

$$f_{X|Y}(x|y) \;\equiv\; \frac{f_{X,Y}(x,y)}{f_Y(y)}\ .$$

When $X$ and $Y$ are *independent* then

$$f_{X|Y}(x|y) \;\equiv\; \frac{f_{X,Y}(x,y)}{f_Y(y)} \;=\; \frac{f_X(x)\ f_Y(y)}{f_Y(y)} \;=\; f_X(x)\ ,$$

(assuming $f_Y(y) \neq 0$ ).

**EXAMPLE** : The random variables with density function

$$f_{X,Y}(x,y) \;=\; \begin{cases} e^{-x-y} & \text{for } x \geq 0 \;\text{ and }\; y \geq 0 \;, \\ \\ 0 & \text{otherwise} \;\;, \end{cases}$$

have (by previous exercise) the marginal density functions

$$f_X(x) \;=\; e^{-x} \quad , \quad f_Y(y) \;=\; e^{-y} \;,$$

for $x \geq 0$ and $y \geq 0$ , and zero otherwise.

Thus for such $x, y$ we have

$$f_{X|Y}(x|y) \;=\; \frac{f_{X,Y}(x,y)}{f_Y(y)} \;=\; \frac{e^{-x-y}}{e^{-y}} \;=\; e^{-x} \;=\; f_X(x) \;,$$

*i.e.*, information about $Y$ does not alter the density function of $X$ .

Indeed, we have already seen that $X$ and $Y$ are *independent* .

## Expectation

The *expected value* of a continuous random variable $X$ is

$$E[X] \quad = \quad \int_{-\infty}^{\infty} x \; f_X(x) \; dx \; ,$$

which represents the *average value* of $X$ over many trials.

The expected value of a *function of a random variable* is

$$E[g(X)] \quad = \quad \int_{-\infty}^{\infty} g(x) \; f_X(x) \; dx \; .$$

The expected value of a function of *two* random variables is

$$E[g(X,Y)] \quad = \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \; f_{X,Y}(x,y) \; dy \; dx \; .$$

**EXAMPLE** :

For the *pointer* experiment

$$f_X(x) \;=\; \begin{cases} 0\,, & x \le 0 \\ 1\,, & 0 < x \le 1 \\ 0\,, & 1 < x \end{cases}$$

we have

$$E[X] \;=\; \int_{-\infty}^{\infty} x\; f_X(x)\; dx \;=\; \int_0^1 x\; dx \;=\; \frac{x^2}{2}\Big|_0^1 \;=\; \frac{1}{2}\,,$$

and

$$E[X^2] \;=\; \int_{-\infty}^{\infty} x^2\; f_X(x)\; dx \;=\; \int_0^1 x^2\; dx \;=\; \frac{x^3}{3}\Big|_0^1 \;=\; \frac{1}{3}\,.$$

**EXAMPLE** : For the joint density function

$$f_{X,Y}(x,y) \;=\; \begin{cases} e^{-x-y} & \text{for } x>0 \;\text{ and }\; y>0\,, \\ \\ 0 & \text{otherwise}\,. \end{cases}$$

we have (by previous exercise) the marginal density functions

$$f_X(x) = \begin{cases} e^{-x} & \text{for } x>0\,, \\ \\ 0 & \text{otherwise}\;\;, \end{cases} \qquad \text{and} \qquad f_Y(y) = \begin{cases} e^{-y} & \text{for } y>0\,, \\ \\ 0 & \text{otherwise}\,. \end{cases}$$

Thus $\quad E[X] \;=\; \displaystyle\int_0^\infty x\, e^{-x}\, dx \;=\; -[(x+1)e^{-x}]\Big|_0^\infty \;=\; 1\,.$ ( Check ! )

Similarly $\qquad E[Y] \;=\; \displaystyle\int_0^\infty y\, e^{-y}\, dy \;=\; 1\,,$

and

$$E[XY] \;=\; \int_0^\infty \int_0^\infty xy\, e^{-x-y}\, dy\, dx \;=\; 1\,. \qquad (\text{ Check ! })$$

**EXERCISE** :

Prove the following for *continuous* random variables :

- $E[aX] \;=\; a\,E[X]\,,$
- $E[aX+b] \;=\; a\,E[X] \;+\; b\,,$
- $E[X+Y] \;=\; E[X] \;+\; E[Y]\,,$

and *compare* the proofs to those for *discrete* random variables.

**EXERCISE** :

A stick of length 1 is split at a randomly selected point $X$.
( Thus $X$ is uniformly distributed in the interval $[0,1]$. )
Determine the expected length of the piece containing the point $1/3$.

**PROPERTY** : If $X$ and $Y$ are *independent* then

$$E[XY] \quad = \quad E[X] \cdot E[Y] \, .$$

**PROOF** :

$$\begin{aligned}
E[XY] &= \int_{\mathbb{R}} \int_{\mathbb{R}} \; x \; y \; f_{X,Y}(x,y) \; \; dy \; dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \; x \; y \; f_X(x) \; f_Y(y) \; \; dy \; dx \qquad \text{(by independence)} \\
&= \int_{\mathbb{R}} [ \; x \; f_X(x) \; \int_{\mathbb{R}} \; y \; f_Y(y) \; dy \; ] \; dx \\
&= [\int_{\mathbb{R}} \; x \; f_X(x) \; dx \; ] \; \cdot \; [\int_{\mathbb{R}} \; y \; f_Y(y) \; dy \; ] \\
&= E[X] \cdot E[Y] \, .
\end{aligned}$$

**REMARK** : Note how the proof parallels that for the discrete case !

**EXAMPLE** : For

$$f_{X,Y}(x,y) \;=\; \begin{cases} e^{-x-y} & \text{for } x>0 \;\text{ and }\; y>0\;, \\ \\ 0 & \text{otherwise}\;, \end{cases}$$

we already found

$$f_X(x) \;=\; e^{-x} \qquad , \qquad f_Y(y) \;=\; e^{-y} \quad ,$$

so that

$$f_{X,Y}(x,y) \;\;=\;\; f_X(x) \;\cdot\; f_Y(y)\;,$$

*i.e.*, $X$ and $Y$ are *independent* .

Indeed, we also already found that

$$E[X] \;\;=\;\; E[Y] \;\;=\;\; E[XY] \;\;=\;\; 1 \quad ,$$

so that

$$E[XY] \;\;=\;\; E[X] \;\cdot\; E[Y]\;.$$

## Variance

Let $$\mu \;=\; E[X] \;=\; \int_{-\infty}^{\infty} x\; f_X(x)\; dx$$

Then the *variance* of the continuous random variable $X$ is

$$Var(X) \;\equiv\; E[\,(X-\mu)^2] \;\equiv\; \int_{-\infty}^{\infty} (x-\mu)^2\; f_X(x)\; dx\;,$$

which is the average weighted *square distance* from the mean.

As in the discrete case, we have

$$\begin{aligned} Var(X) \;&=\; E[X^2 - 2\mu X \;+\; \mu^2] \\ &=\; E[X^2] - 2\mu E[X] \;+\; \mu^2 \;=\; E[X^2] \;-\; \mu^2\;. \end{aligned}$$

The *standard deviation* of $X$ is

$$\sigma(X) \;\equiv\; \sqrt{Var(X)} \;=\; \sqrt{E[X^2] \;-\; \mu^2}\;.$$

which is the average weighted *distance* from the mean.

**EXAMPLE** : For $f(x) = \begin{cases} e^{-x} , & x > 0 , \\ 0 , & x \leq 0 , \end{cases}$

we have

$$E[X] = \mu = \int_0^\infty x\, e^{-x}\, dx = 1 \quad \text{(already done!)} ,$$

$$E[X^2] = \int_0^\infty x^2\, e^{-x}\, dx = -\left[(x^2 + 2x + 2)e^{-x}\right]\Big|_0^\infty = 2 ,$$

$$Var(X) = E[X^2] - \mu^2 = 2 - 1^2 = 1 ,$$

$$\sigma(X) = \sqrt{Var(X)} = 1 .$$

**NOTE** : The two integrals can be done by "*integration by parts* ".

**EXERCISE** :

Also use the *Method of Moments* to compute $E[X]$ and $E[X^2]$ .

**EXERCISE** : For the random variable $X$ with density function

$$f(x) = \begin{cases} 0\,, & x \leq -1 \\ c\,, & -1 < x \leq 1 \\ 0\,, & x > 1 \end{cases}$$

- Determine the value of $c$
- Draw the graph of $f(x)$
- Determine the distribution function $F(x)$
- Draw the graph of $F(x)$
- Determine $E[X]$
- Compute $Var(X)$ and $\sigma(X)$
- Determine $P(X \leq -\frac{1}{2})$
- Determine $P(\mid X \mid \geq \frac{1}{2})$

**EXERCISE** : For the random variable $X$ with density function

$$f(x) \quad = \quad \begin{cases} x+1\ , & -1 < x \leq 0 \\ 1-x\ , & 0 < x \leq 1 \\ 0\ , & \text{otherwise} \end{cases}$$

- Draw the graph of $f(x)$
- Verify that $\int_{-\infty}^{\infty} f(x)\ dx = 1$
- Determine the distribution function $F(x)$
- Draw the graph of $F(x)$
- Determine $E[X]$
- Compute $Var(X)$ and $\sigma(X)$
- Determine $P(X \geq \frac{1}{3})$
- Determine $P(\mid X \mid \leq \frac{1}{3})$

**EXERCISE** : For the random variable $X$ with density function

$$f(x) = \begin{cases} \frac{3}{4}\,(1-x^2)\,, & -1 < x \leq 1 \\ 0\,, & \text{otherwise} \end{cases}$$

- Draw the graph of $f(x)$
- Verify that $\int_{-\infty}^{\infty} f(x)\,dx = 1$
- Determine the distribution function $F(x)$
- Draw the graph of $F(x)$
- Determine $E[X]$
- Compute $Var(X)$ and $\sigma(X)$
- Determine $P(X \leq 0)$
- Compute $P(X \geq \frac{2}{3})$
- Compute $P(\mid X \mid \geq \frac{2}{3})$

**EXERCISE** : Recall the density function

$$f_n(x) = \begin{cases} cx^n(1-x^n)\ , & 0 \le x \le 1 \\ \\ 0\ , & \text{otherwise} \end{cases}$$

considered earlier, where $n$ is a positive integer, and where

$$c \ = \ \frac{(n+1)(2n+1)}{n} \ .$$

- Determine $E[X]$ .
- What happens to $E[X]$ for *large* $n$ ?
- Determine $E[X^2]$
- What happens to $E[X^2]$ for *large* $n$ ?
- What happens to $Var(X)$ for *large* $n$ ?

## Covariance

Let $X$ and $Y$ be continuous random variables with *mean*

$$E[X] \;=\; \mu_X \qquad , \qquad E[Y] \;=\; \mu_Y \;.$$

Then the *covariance* of $X$ and $Y$ is

$$Cov(X,Y) \;\equiv\; E[\,(X-\mu_X)\,(Y-\mu_Y)\,]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)\,(y-\mu_Y)\,f_{X,Y}(x,y)\,dy\,dx \;.$$

As in the discrete case, we have

$$\begin{aligned} Cov(X,Y) &= E[\,(X-\mu_X)\,(Y-\mu_Y)\,] \\ &= E[XY - \mu_X Y - \mu_Y X \;+\; \mu_X\mu_Y] \\ &= E[XY] \;-\; E[X]\;E[Y] \;. \end{aligned}$$

As in the discrete case, we also have

**PROPERTY 1** :

- $Var(X+Y) \;=\; Var(X) \;+\; Var(Y) \;+\; 2\,Cov(X,Y)$ ,

and

**PROPERTY 2** : If $X$ and $Y$ are *independent* then

- $Cov(X,Y) \;=\; 0$ ,

- $Var(X+Y) \;=\; Var(X) \;+\; Var(Y)$ .

**NOTE** :

- The proofs are identical to those for the discrete case !

- As in the discrete case, if $Cov(X,Y) \;=\; 0$ then $X$ and $Y$ are not necessarily independent!

**EXAMPLE** : For

$$f_{X,Y}(x,y) \;=\; \begin{cases} e^{-x-y} & \text{for } x>0 \;\text{ and }\; y>0\;, \\ \\ 0 & \text{otherwise}\;, \end{cases}$$

we already found

$$f_X(x) \;=\; e^{-x} \qquad , \qquad f_Y(y) \;=\; e^{-y} \quad ,$$

so that

$$f_{X,Y}(x,y) \;=\; f_X(x)\;\cdot\; f_Y(y)\;,$$

*i.e.*, $X$ and $Y$ are *independent* .

Indeed, we also already found

$$E[X] \;=\; E[Y] \;=\; E[XY] \;=\; 1 \quad ,$$

so that

$$Cov(X,Y) \;=\; E[XY] \;-\; E[X]\;E[Y] \;=\; 0\;.$$

**EXERCISE** :

Verify the following properties :

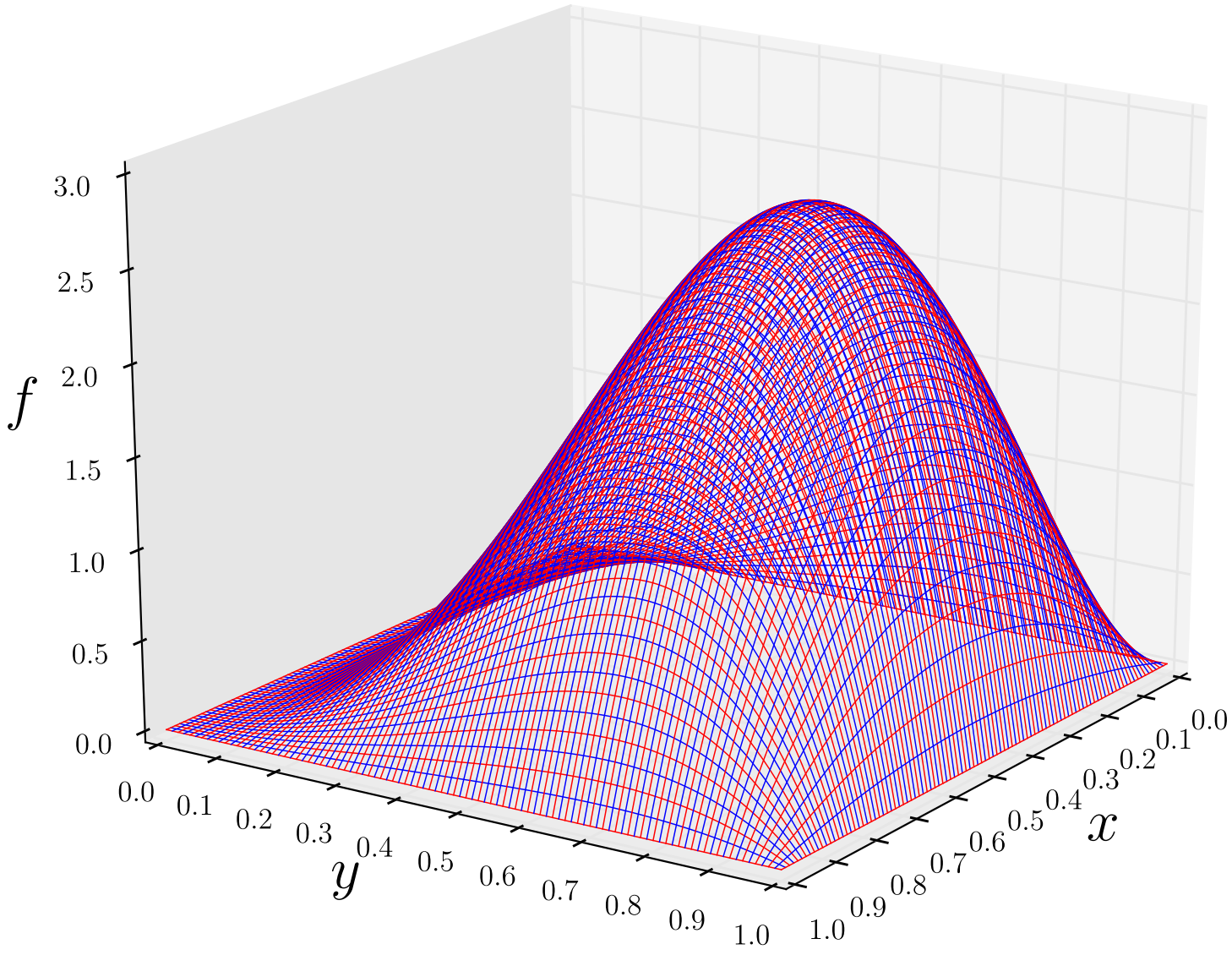
- $Var(cX + d) \;=\; c^2\, Var(X)\;,$
- $Cov(X, Y) \;=\; Cov(Y, X)\;,$
- $Cov(cX, Y) \;=\; c\, Cov(X, Y)\;,$
- $Cov(X, cY) \;=\; c\, Cov(X, Y)\;,$
- $Cov(X + Y, Z) \;=\; Cov(X, Z) \;+\; Cov(Y, Z)\;,$
- $Var(X + Y) \;=\; Var(X) \;+\; Var(Y) \;+\; 2\, Cov(X, Y)\;.$

**EXERCISE** :

For the random variables $X$ , $Y$ with *joint density function*

$$f(x,y) \;=\; \begin{cases} 45xy^2(1-x)(1-y^2)\ , & 0 \le x \le 1\ ,\ 0 \le y \le 1 \\ \\ 0\ , & \text{otherwise} \end{cases}$$

- Verify that $\int_0^1 \int_0^1 f(x,y)\ dy\ dx = 1$ .
- Determine the *marginal density functions* $f_X(x)$ and $f_Y(y)$ .
- Are $X$ and $Y$ *independent* ?
- What is the value of $Cov(X,Y)$ ?




The joint probability density function $f_{XY}(x, y)$ .

**Markov's inequality**.

For a continuous *nonnegative* random variable $X$, and $c > 0$, we have

$$P(X \geq c) \quad \leq \quad \frac{E[X]}{c} \; .$$

**PROOF** :

$$\begin{aligned} E[X] \quad = \quad \int_0^\infty x f(x)\, dx \quad &= \quad \int_0^c x f(x)\, dx \quad + \quad \int_c^\infty x f(x)\, dx \\ &\geq \quad \int_c^\infty x f(x)\, dx \\ &\geq \quad c \int_c^\infty f(x)\, dx \qquad \text{( Why ? )} \\ &= \quad c\, P(X \geq c) \; . \end{aligned}$$

**EXERCISE** :

Show Markov's inequality also holds for *discrete* random variables.

**Markov's inequality** : For continuous *nonnegative* $X$ , $c > 0$ :

$$P(X \geq c) \quad \leq \quad \frac{E[X]}{c} \ .$$

**EXAMPLE** : For

$$f(x) \ = \ \begin{cases} e^{-x} & \text{for } x > 0 \ , \\ \\ 0 & \text{otherwise} \ , \end{cases}$$

we have

$$E[X] \quad = \quad \int_0^\infty x \ e^{-x} \ dx \quad = \quad 1 \qquad \text{(already done!)}$$

Markov's inequality gives

$$c = \mathbf{1} \ : \qquad P(X \geq \mathbf{1}) \quad \leq \quad \frac{E[X]}{\mathbf{1}} \quad = \quad \frac{1}{\mathbf{1}} \quad = \quad 1 \ \ (!)$$

$$c = \mathbf{10} \ : \qquad P(X \geq \mathbf{10}) \quad \leq \quad \frac{E[X]}{\mathbf{10}} \quad = \quad \frac{1}{10} \quad = \quad 0.1$$

**QUESTION** : Are these estimates "*sharp* " ?

**QUESTION** : Are these estimates "*sharp*" ?

Markov's inequality gives

$$c = 1 \ : \qquad P(X \geq 1) \quad \leq \quad \frac{E[X]}{1} \quad = \quad \frac{1}{1} \quad = \quad 1 \ \ (!)$$

$$c = 10 \ : \qquad P(X \geq 10) \quad \leq \quad \frac{E[X]}{10} \quad = \quad \frac{1}{10} \quad = \quad 0.1$$

The actual values are

$$P(X \geq 1) \quad = \quad \int_1^\infty e^{-x} \, dx \quad = \quad e^{-1} \quad \cong \ 0.37$$

$$P(X \geq 10) \quad = \quad \int_{10}^\infty e^{-x} \, dx \quad = \quad e^{-10} \quad \cong \ 0.000045$$

**EXERCISE** : Suppose the score of students taking an examination is a random variable with mean 65 .

Give an upper bound on the probability that a student's score is greater than 75.

**Chebyshev's inequality**: For (practically) any random variable $X$:

$$P(\ |X-\mu|\ \geq\ k\,\sigma\ )\ \leq\ \frac{1}{k^2}\ ,$$

where $\mu = E[X]$ is the *mean*, $\sigma = \sqrt{Var(X)}$ the *standard deviation*.

**PROOF** : Let $Y \equiv (X-\mu)^2$ , which is nonnegative.

By Markov's inequality

$$P(Y \geq c)\ \leq\ \frac{E[Y]}{c}\ .$$

Taking $c = k^2\sigma^2$ we have

$$P(\ |X-\mu|\ \geq\ k\sigma\ )\ =\ P(\ (X-\mu)^2\ \geq\ k^2\sigma^2\ )\ =\ P(\ Y\ \geq\ k^2\sigma^2\ )$$

$$\leq\ \frac{E[\ Y\ ]}{k^2\sigma^2}\ =\ \frac{Var(X)}{k^2\sigma^2}\ =\ \frac{\sigma^2}{k^2\sigma^2}\ =\ \frac{1}{k^2}\ .\quad \textbf{QED}\ !$$

**NOTE** : This inequality also holds for discrete random variables.

**EXAMPLE** : Suppose the value of the Canadian dollar in terms of the US dollar over a certain period is a random variable $X$ with

$$\textit{mean} \quad \mu = 0.98 \quad \text{and} \quad \textit{standard deviation} \quad \sigma = 0.05 \ .$$

What can be said of the probability that the Canadian dollar is valued

$$\text{between} \quad \$0.88\text{US} \quad \text{and} \quad \$1.08\text{US} \ ,$$

that is,

$$\text{between} \quad \mu - 2\sigma \quad \text{and} \quad \mu + 2\sigma \ ?$$

**SOLUTION** : By Chebyshev's inequality we have

$$P(\ |X - \mu| \ \geq \ 2\,\sigma\ ) \ \leq \ \frac{1}{2^2} \ = \ 0.25 \ .$$

Thus

$$P(\ |X - \mu| \ < \ 2\,\sigma\ ) \ > \ 1 \ - \ 0.25 \ = \ 0.75 \ ,$$

that is,

$$P(\ \$0.88\text{US} \ < \ \text{Can}\$ \ < \ \$1.08\text{US}\ ) \ > \ 75\ \% \ .$$

**EXERCISE** :

The score of students taking an examination is a random variable with mean $\mu = 65$ and standard deviation $\sigma = 5$ .

- What is the probability a student scores between 55 and 75 ?
- How many students would have to take the examination so that the probability that their average grade is between 60 and 70 is at least 80% ?

  **HINT** : Defining

$$\bar{X} \;=\; \frac{1}{n}X_1 + \frac{1}{n}X_2 + \;\cdots\; + \frac{1}{n}X_n \;, \quad (\text{ the average grade })$$

  we have

$$\mu_{\bar{X}} \;=\; E[\bar{X}] \;=\; n \cdot \frac{1}{n}\,\mu \;=\; \mu \;=\; 65 \;,$$

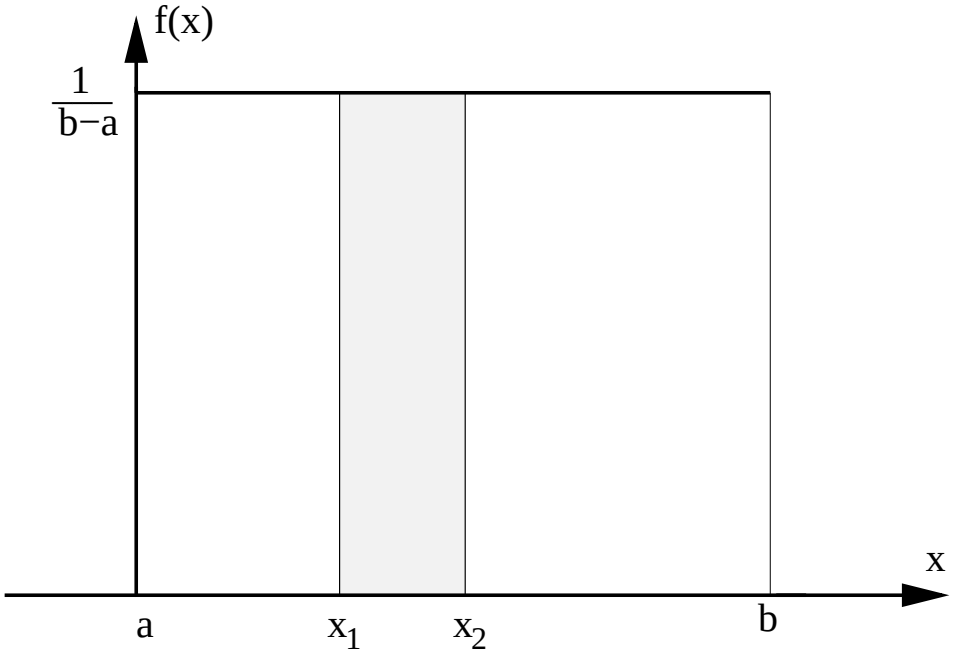
  and, assuming independence,

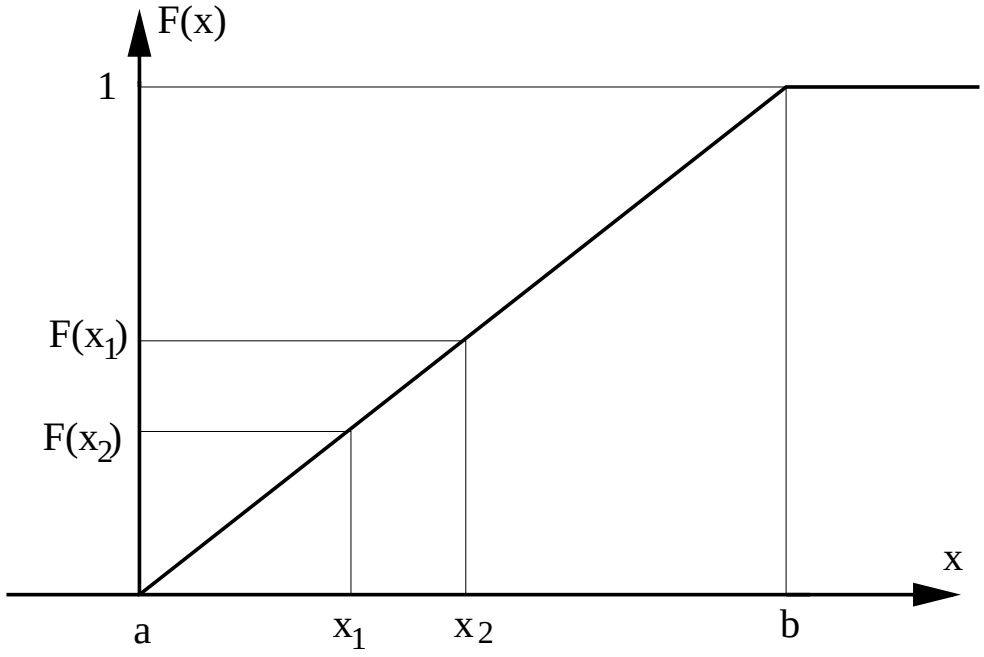
$$Var(\bar{X}) \;=\; n\,\frac{\sigma^2}{n^2} \;=\; \frac{\sigma^2}{n} \;=\; \frac{25}{n} \;, \quad \text{and} \quad \sigma_{\bar{X}} \;=\; \frac{5}{\sqrt{n}} \;.$$

# SPECIAL CONTINUOUS RANDOM VARIABLES

## The Uniform Random Variable

$$f(x) = \begin{cases} \frac{1}{b-a}\ , & a < x \leq b \\ \\ 0\ , & \text{otherwise} \end{cases} \qquad , \qquad F(x) = \begin{cases} 0\ , & x \leq a \\ \frac{x-a}{b-a}\ , & a < x \leq b \\ 1\ , & x > b \end{cases}$$




(Already introduced earlier for the special case $a = 0$, $b = 1$ .)

**EXERCISE** :

Show that the *uniform density function*

$$f(x) = \begin{cases} \frac{1}{b-a} \ , & a < x \leq b \\ \\ 0 \ , & \text{otherwise} \end{cases}$$

has *mean*

$$\mu \quad = \quad \frac{a+b}{2} \ ,$$

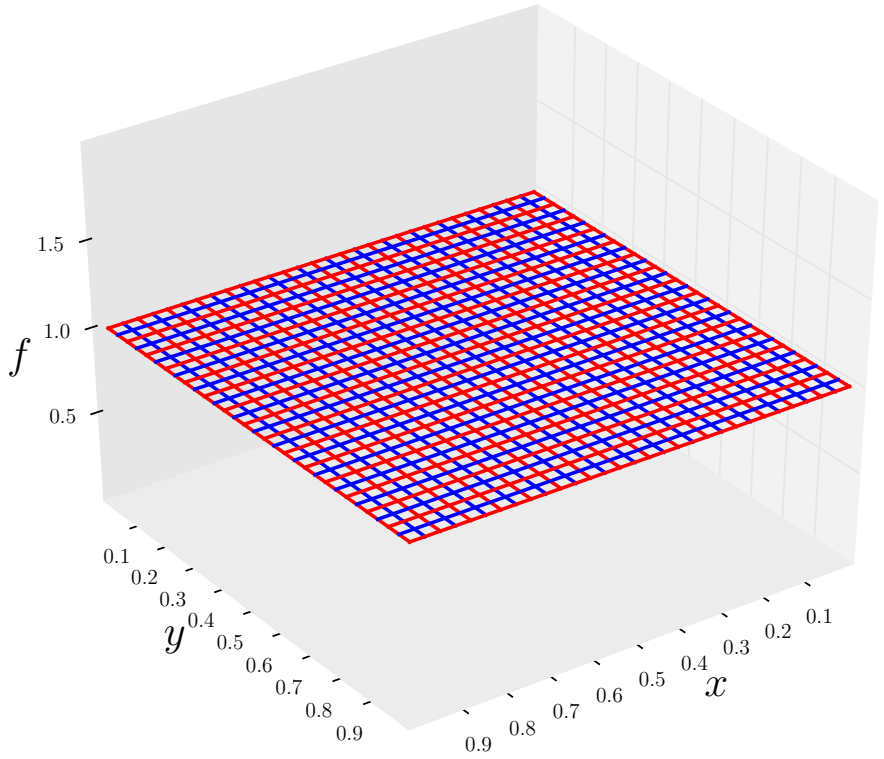
and *standard deviation*

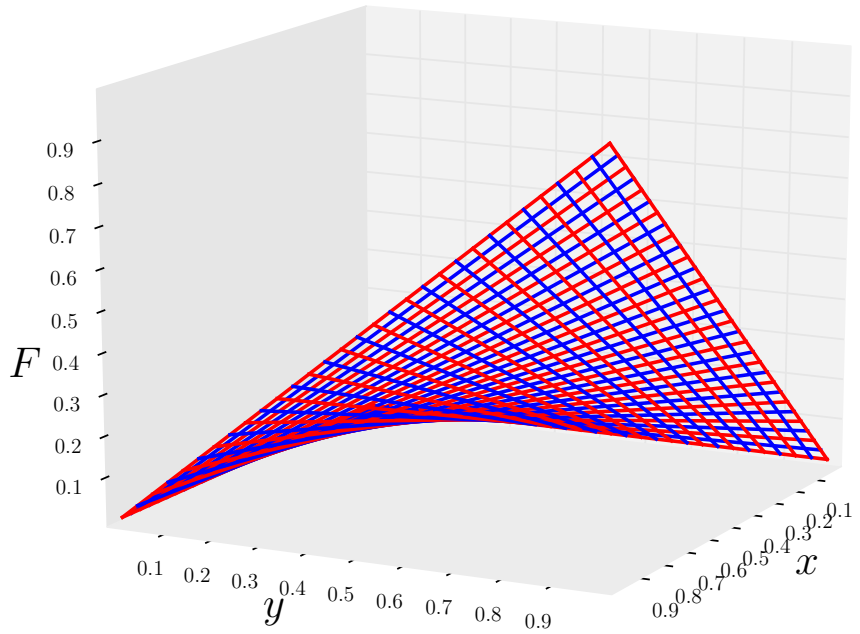
$$\sigma \quad = \quad \frac{b-a}{2\sqrt{3}} \ .$$

A *joint uniform* random variable :

$$f(x,y) \;=\; \frac{1}{(b-a)(d-c)} \quad , \quad F(x,y) \;=\; \frac{(x-a)(y-c)}{(b-a)(d-c)} \;,$$

for $x \in (a,b],\ y \in (c,d]$.




Here $x \in [0,1],\ y \in [0,1]$ .

**EXERCISE** :

Consider the *joint uniform density function*

$$f(x,y) \;=\; \begin{cases} c & \text{for } \; x^2+y^2 \;\leq\; 4\;, \\ \\ 0 & \text{otherwise}\;. \end{cases}$$

- What is the value of $c$ ?
- What is $P(X<0)$ ?
- What is $P(X<0\;,\;Y<0)$ ?
- What is $f(\;x \mid y=1\;)$ ?

**HINT** : No complicated calculations are needed !

## The Exponential Random Variable

$$f(x) = \begin{cases} \lambda e^{-\lambda x} , & x > 0 \\ \\ 0 , & x \leq 0 \end{cases} \quad , \quad F(x) = \begin{cases} 1 - e^{-\lambda x} , & x > 0 \\ \\ 0 , & x \leq 0 \end{cases}$$

with

$$E[X] = \mu = \int_0^\infty x \, \lambda e^{-\lambda x} \, dx = \frac{1}{\lambda} \quad (\text{ Check ! }) ,$$

$$E[X^2] = \int_0^\infty x^2 \, \lambda e^{-\lambda x} \, dx = \frac{2}{\lambda^2} \quad (\text{ Check ! }) ,$$

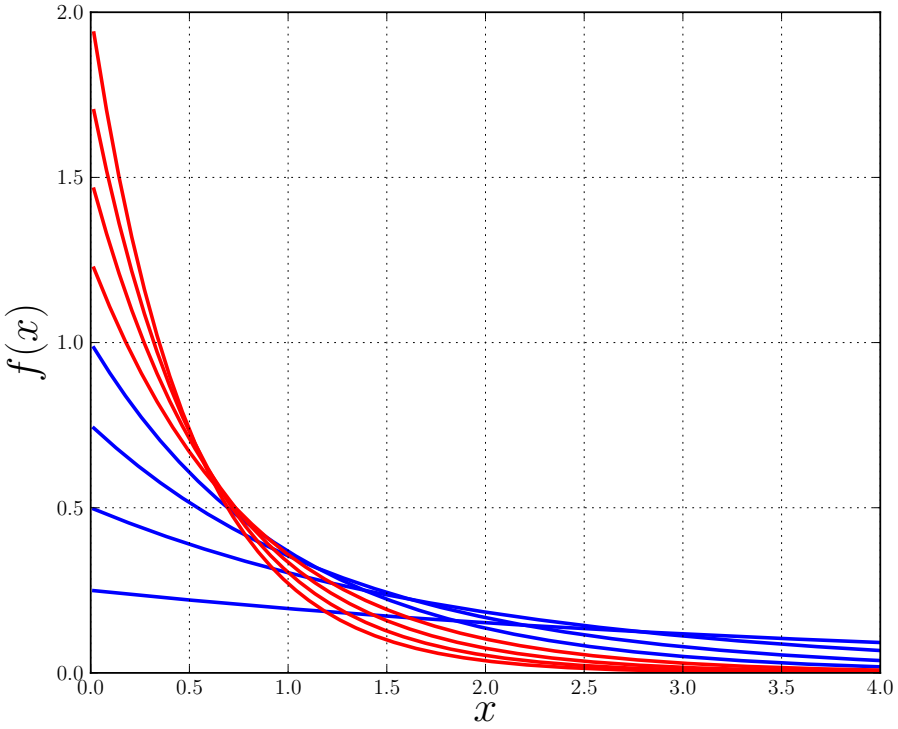
$$Var(X) = E[X^2] - \mu^2 = \frac{1}{\lambda^2} ,$$

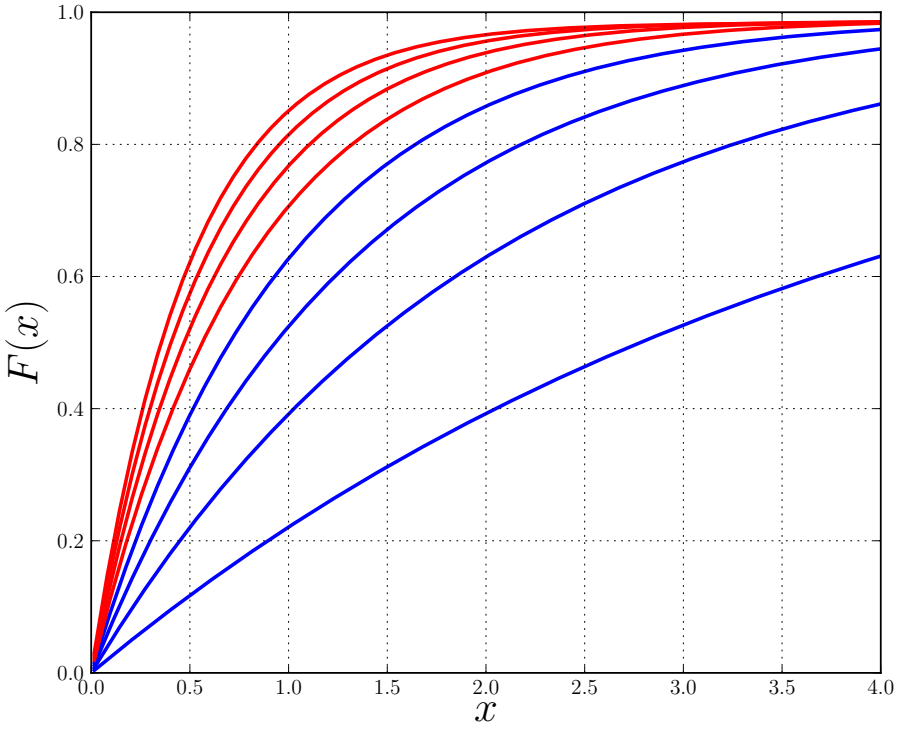
$$\sigma(X) = \sqrt{Var(X)} = \frac{1}{\lambda} .$$

**NOTE** : The two integrals can be done by "*integration by parts* ".

**EXERCISE** : (Done earlier for $\lambda = 1$) :

Also use the *Method of Moments* to compute $E[X]$ and $E[X^2]$ .




The Exponential *density* and *distribution* functions

$$f(x) \;=\; \lambda e^{-\lambda x} \qquad , \qquad F(x) \;=\; 1 - e^{-\lambda x} \;,$$

for $\lambda = 0.25, 0.50, 0.75, 1.00$ (*blue*), $1.25, 1.50, 1.75, 2.00$ (*red* ).

**PROPERTY** : From

$$F(x) \equiv P(X \leq x) = 1 - e^{-\lambda x} , \qquad (\text{for } x > 0 ) ,$$

we have $$P(X > x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x} .$$

Also, for $\Delta x > 0$ ,

$$P(X > x + \Delta x \mid X > x) = \frac{P(X > x + \Delta x \ , \ X > x)}{P(X > x)}$$

$$= \frac{P(X > x + \Delta x)}{P(X > x)} = \frac{e^{-\lambda(x+\Delta x)}}{e^{-\lambda x}} = e^{-\lambda \Delta x} .$$

**CONCLUSION** : $P( \ X > x + \Delta x \ \mid \ X > x \ )$

*only depends* on $\Delta x$ (and $\lambda$), and *not* on $x$ !

We say that the exponential random variable is "*memoryless*" .

**EXAMPLE** :

Let the density function $f(t)$ model *failure* of a device,

$$f(t) \quad = \quad e^{-t} \, , \qquad \text{(taking } \lambda = 1 \text{ )} \, ,$$

*i.e.*, the *probability of failure* in the time-interval $(0, t]$ is

$$F(t) \quad = \quad \int_0^t f(t) \, dt \quad = \quad \int_0^t e^{-t} \, dt \quad = \quad 1 \, - \, e^{-t} \, ,$$

with

$$F(0) \; = \; 0 \, , \qquad \text{( the device works at time 0 ).}$$

and

$$F(\infty) \; = \; 1 \, , \qquad \text{( the device must fail at some time ).}$$

**EXAMPLE** : ( continued $\cdots$ ) $\qquad F(t) \;=\; 1 \,-\, e^{-t}$ .

Let $E_t$ be the *event* that the device still *works* at time $t$ :

$$P(E_t) \;=\; 1 \,-\, F(t) \;=\; e^{-t} \,.$$

The probability it still works at time $t+1$ is

$$P(E_{t+1}) \;=\; 1 \,-\, F(t+1) \;=\; e^{-(t+1)} \,.$$

The probability it still works at time $t+1$, given it works at time $t$ is

$$P(E_{t+1}|E_t) \;=\; \frac{P(E_{t+1}E_t)}{P(E_t)} \;=\; \frac{P(E_{t+1})}{P(E_t)} \;=\; \frac{e^{-(t+1)}}{e^{-t}} \;=\; \frac{1}{e} \,,$$

which is *independent of* $t$ !

**QUESTION** : Is such an exponential distribution realistic if the "device" is your heart, and time $t$ is measured in decades ?

## The Standard Normal Random Variable

The *standard normal* random variable has *density function*

$$f(x) \quad = \quad \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2} \;, \qquad\qquad -\infty < x < \infty \;,$$

with *mean*

$$\mu \quad = \quad \int_{-\infty}^{\infty} x\; f(x)\; dx \quad = \quad 0 \;, \qquad\qquad (\text{ Check ! })$$

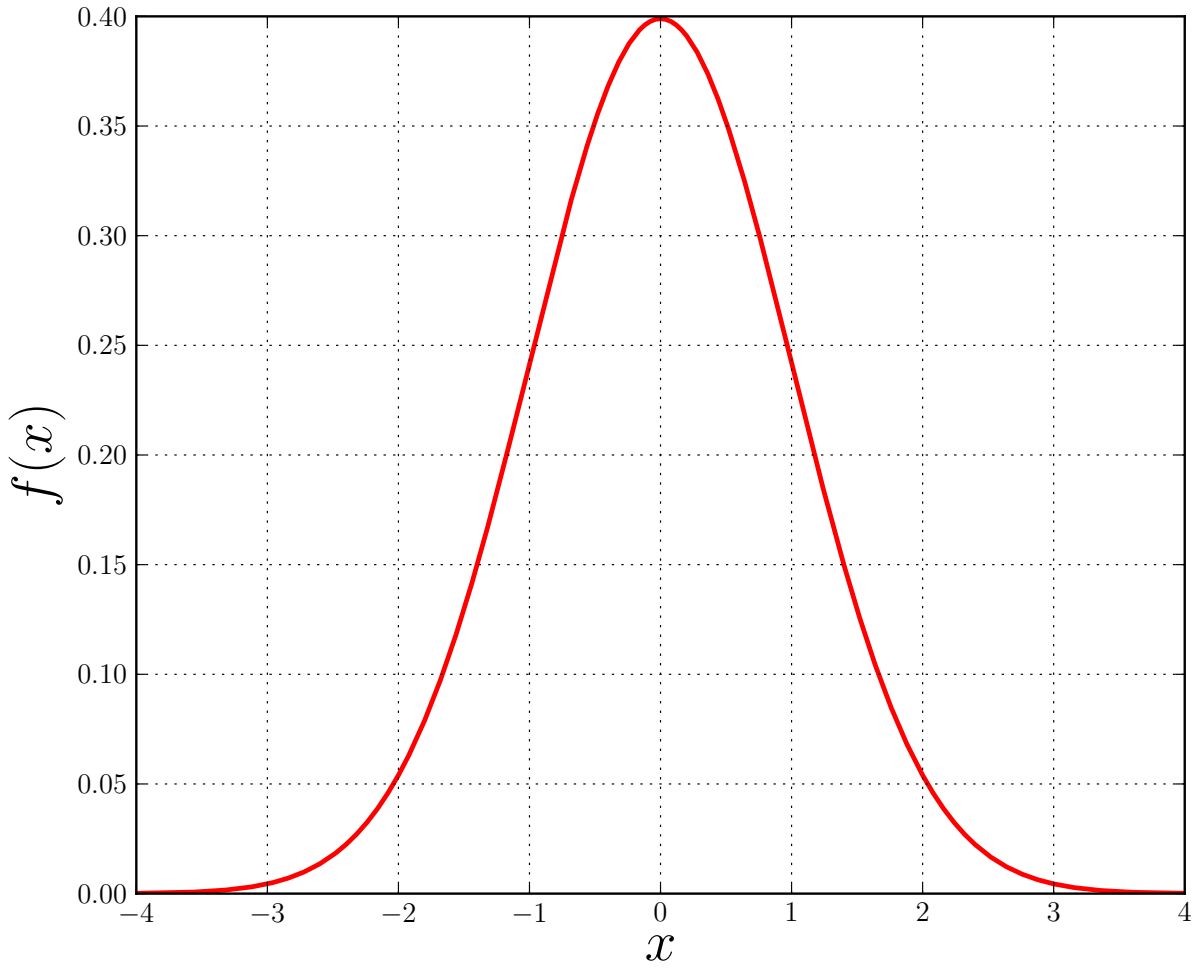
Since

$$E[X^2] \quad = \quad \int_{-\infty}^{\infty} x^2\; f(x)\; dx \quad = \quad 1 \;, \qquad\qquad (\text{ more difficult } \cdots)$$

we have

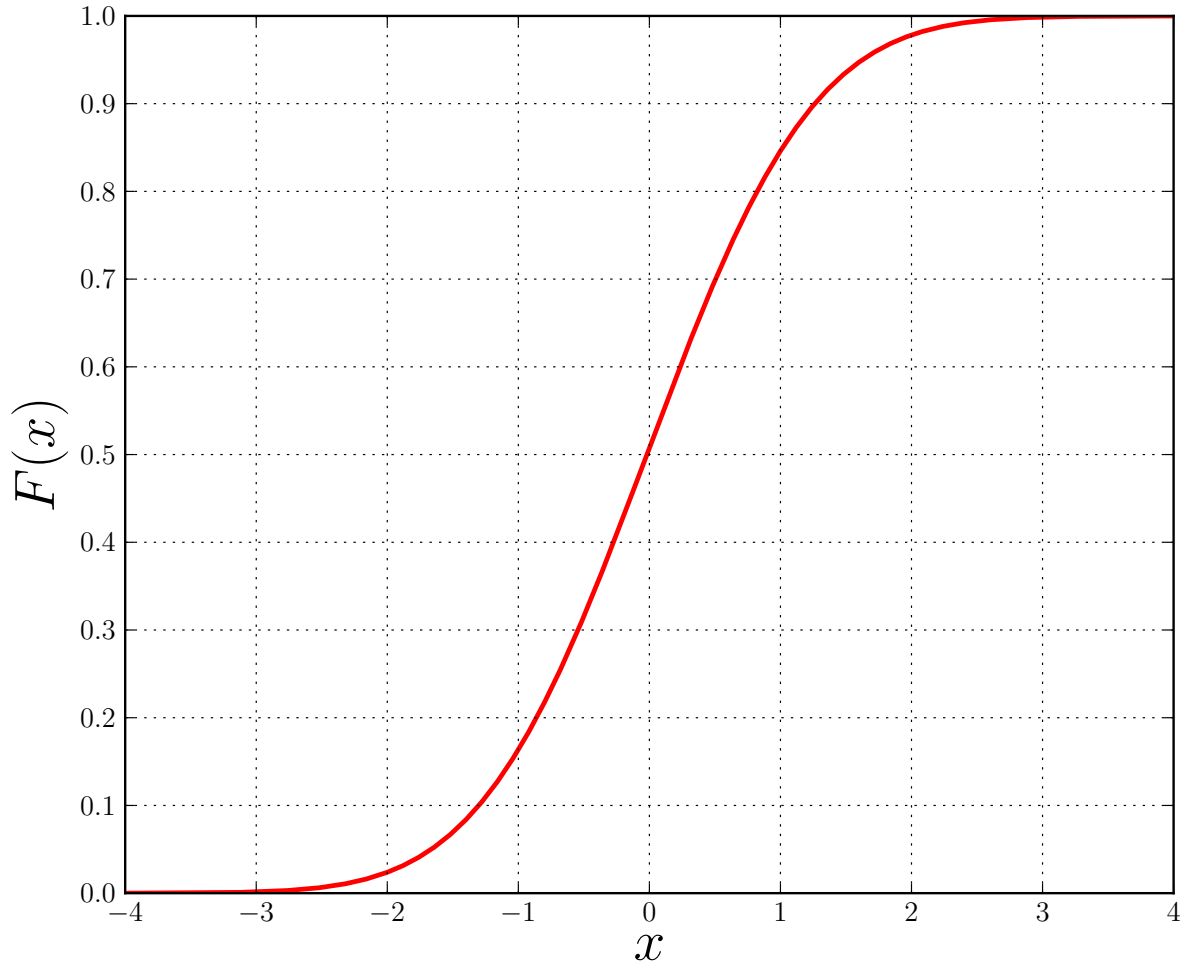
$$Var(X) \;=\; E[X^2] - \mu^2 \quad = \quad 1 \;, \qquad \text{and} \qquad \sigma(X) \;=\; 1 \;.$$

$$f(x) \quad = \quad \frac{1}{\sqrt{2\pi}} \, e^{-\frac{1}{2}x^2}$$




The *standard normal density function* $f(x)$ .

$$\mathbf{\Phi(x)} \quad = \quad F(x) \quad = \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}x^2} \, dx$$




The *standard normal distribution function* $F(x)$

( often denoted by $\mathbf{\Phi(x)}$ ) .

## The Standard Normal Distribution $\Phi(z)$

| $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ |
|---|---|---|---|
| 0.0 | .5000 | -1.2 | .1151 |
| -0.1 | .4602 | -1.4 | .0808 |
| -0.2 | .4207 | -1.6 | .0548 |
| -0.3 | .3821 | -1.8 | .0359 |
| -0.4 | .3446 | **-2.0** | **.0228** |
| -0.5 | .3085 | -2.2 | .0139 |
| -0.6 | .2743 | -2.4 | .0082 |
| -0.7 | .2420 | -2.6 | .0047 |
| -0.8 | .2119 | -2.8 | .0026 |
| -0.9 | .1841 | -3.0 | .0013 |
| -1.0 | .1587 | -3.2 | .0007 |

( For example, $P(Z \leq -2.0) = \Phi(\mathbf{-2.0}) = \mathbf{2.28}\%$ )

**QUESTION** : How to get the values of $\Phi(z)$ for *positive* $z$ ?

**EXERCISE** :

Suppose the random variable $X$ has the *standard normal* distribution.

What are the values of

- $P(\ X\ \leq -0.5\ )$
- $P(\ X\ \leq\ 0.5\ )$
- $P(\ |X|\ \geq\ 0.5\ )$
- $P(\ |X|\ \leq\ 0.5\ )$
- $P(\ -1\ \leq\ X\ \leq\ 1\ )$
- $P(\ -1\ \leq\ X\ \leq\ 0.5\ )$

## The General Normal Random Variable

The *general normal density function* is

$$f(x) \quad = \quad \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

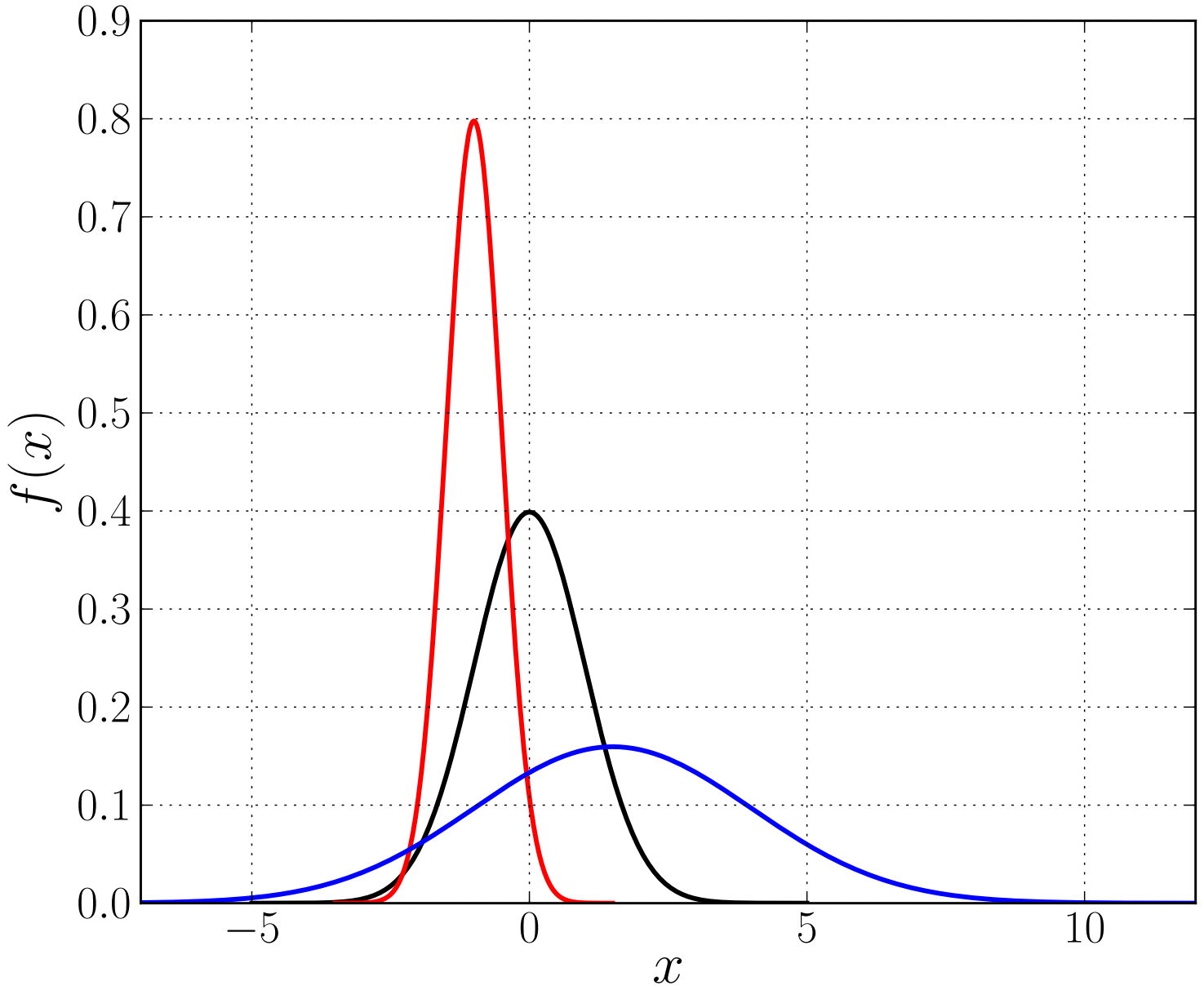
where, not surprisingly,

$$E[X] \quad = \quad \mu \qquad\qquad (\text{ Why ? })$$

One can also show that

$$Var(X) \quad \equiv \quad E[(X-\mu)^2] \quad = \quad \sigma^2\ ,$$

so that $\sigma$ is in fact the *standard deviation* .




The standard normal (*black*) and the normal density functions with $\mu = -1$, $\sigma = 0.5$ (*red* ) and $\mu = 1.5$, $\sigma = 2.5$ (*blue*).

To compute the *mean* of the *general normal density function*

$$f(x) \quad = \quad \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

consider

$$\begin{aligned} E[X-\mu] \quad &= \quad \int_{-\infty}^{\infty}(x-\mu)\ f(x)\ dx \\ &= \quad \frac{1}{\sqrt{2\pi}\ \sigma}\ \int_{-\infty}^{\infty}(x-\mu)\ e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}\ \ dx \\ &= \quad \frac{-\sigma^2}{\sqrt{2\pi}\ \sigma}\ \ e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}\ \ \Big|_{-\infty}^{\infty} \quad = \quad 0\ . \end{aligned}$$

Thus the *mean* is indeed

$$E[X] \quad = \quad \mu\ .$$

**NOTE** : If $X$ is *general normal* we have the *very useful formula* :

$$P(\frac{X-\mu}{\sigma} \leq c) = \Phi(c) ,$$

*i.e.*, we can use the *Table* of the *standard normal distribution* !

**PROOF** : For any constant $c$ we have

$$P(\frac{X-\mu}{\sigma} \leq c) = P(X \leq \mu+c\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\mu+c\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}\, dx .$$

Let $y \equiv (x-\mu)/\sigma$ , so that $x = \mu + y\sigma$ .

Then the new variable $y$ ranges from $-\infty$ to $c$ , and

$$(x-\mu)^2/\sigma^2 = y^2 \quad , \quad dx = \sigma\, dy ,$$

so that

$$P(\frac{X-\mu}{\sigma} \leq c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{-\frac{1}{2}y^2}\, dy = \Phi(c) .$$

( the *standard normal distribution* )

**EXERCISE** : Suppose $X$ is normally distributed with

$$\textit{mean} \quad \mu = 1.5 \qquad \text{and} \qquad \textit{standard deviation} \quad \sigma = 2.5 \ .$$

Use the *standard normal Table* to determine :

- $P(\ X \ \leq -0.5\ )$
- $P(\ X \ \geq \ 0.5\ )$
- $P(\ \ |\, X - \mu \,| \ \ \geq \ 0.5\ )$
- $P(\ \ |\, X - \mu \,| \ \ \leq \ 0.5\ )$

## The Chi-Square Random Variable

Suppose $X_1 \ , \ X_2 \ , \ \cdots \ , \ X_n \ ,$

are *independent standard normal* random variables.

Then $\chi^2_n \equiv X_1^2 + X_2^2 + \cdots + X_n^2 \ ,$

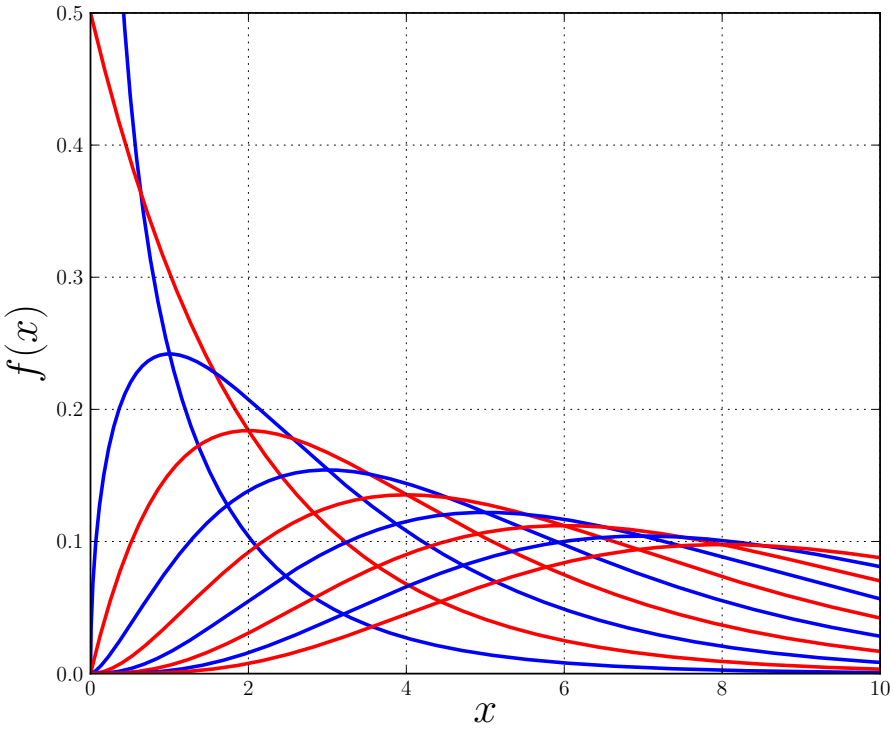
is called the *chi-square random variable* with $n$ *degrees of freedom.*

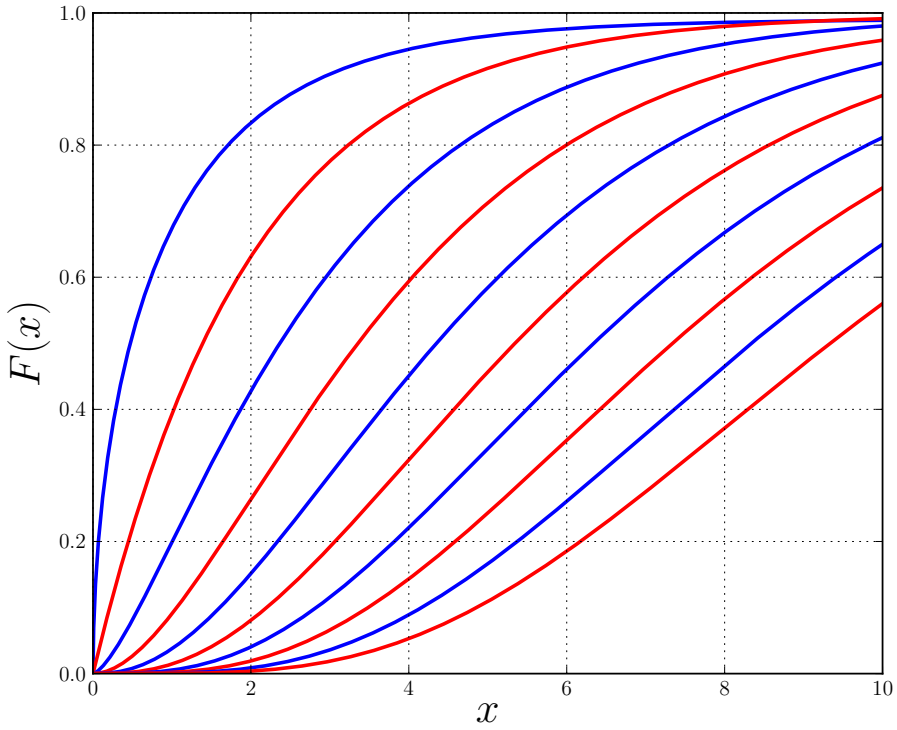
We will show that

$$E[\chi_n^2] = n \quad , \quad Var(\chi_n^2) = 2n \quad , \quad \sigma(\chi_n^2) = \sqrt{2n} \ .$$

**NOTE** :

The $^2$ in $\chi_n^2$ is part of its *name* , while $^2$ in $X_1^2$, *etc.* is "*power* 2 " !




The Chi-Square *density* and *distribution* functions for $n = 1, 2, \cdots, 10$.

( In the Figure for $F$, the value of $n$ increases from left to right. )

If $n = 1$ then

$$\chi_1^2 \quad \equiv \quad X_1^2 \,, \quad \text{where} \quad X \equiv X_1 \quad \text{is } \textit{standard normal} \,.$$

We can compute the *moment generating function* of $\chi_1^2$ :

$$E[e^{t\chi_1^2}] \;=\; E[e^{tX^2}] \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-\frac{1}{2}x^2} \, dx$$

$$= \; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2(1-2t)} \, dx$$

Let

$$1 - 2t \;=\; \frac{1}{\hat{\sigma}^2} \quad , \quad \text{or equivalently} \quad , \quad \hat{\sigma} \;\equiv\; \frac{1}{\sqrt{1-2t}} \,.$$

Then

$$E[e^{t\chi_1^2}] \;=\; \hat{\sigma} \cdot \frac{1}{\sqrt{2\pi}\;\hat{\sigma}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2/\hat{\sigma}^2} \, dx \;=\; \hat{\sigma} \;=\; \frac{1}{\sqrt{1-2t}} \,.$$

(integral of a normal density function)

Thus we have found that :

The *moment generating function* of $\chi^2_1$ is

$$\psi(t) \equiv E[e^{t\chi^2_1}] = \frac{1}{\sqrt{1-2t}} ,$$

with which we can compute

$$E[\chi^2_1] = \psi'(0) = 1 , \quad (\text{ Check ! })$$

$$E[(\chi^2_1)^2] = \psi''(0) = 3 , \quad (\text{ Check ! })$$

$$Var(\chi^2_1) = E[(\chi^2_1)^2] - E[\chi^2_1]^2 = 2 .$$

We found that

$$E[\chi_1^2] \;=\; 1 \qquad , \qquad Var(\chi_1^2) \;=\; 2\,.$$

In the *general case* where

$$\chi_n^2 \;\equiv\; X_1^2 \,+\, X_2^2 \,+\, \cdots \,+\, X_n^2\,,$$

we have

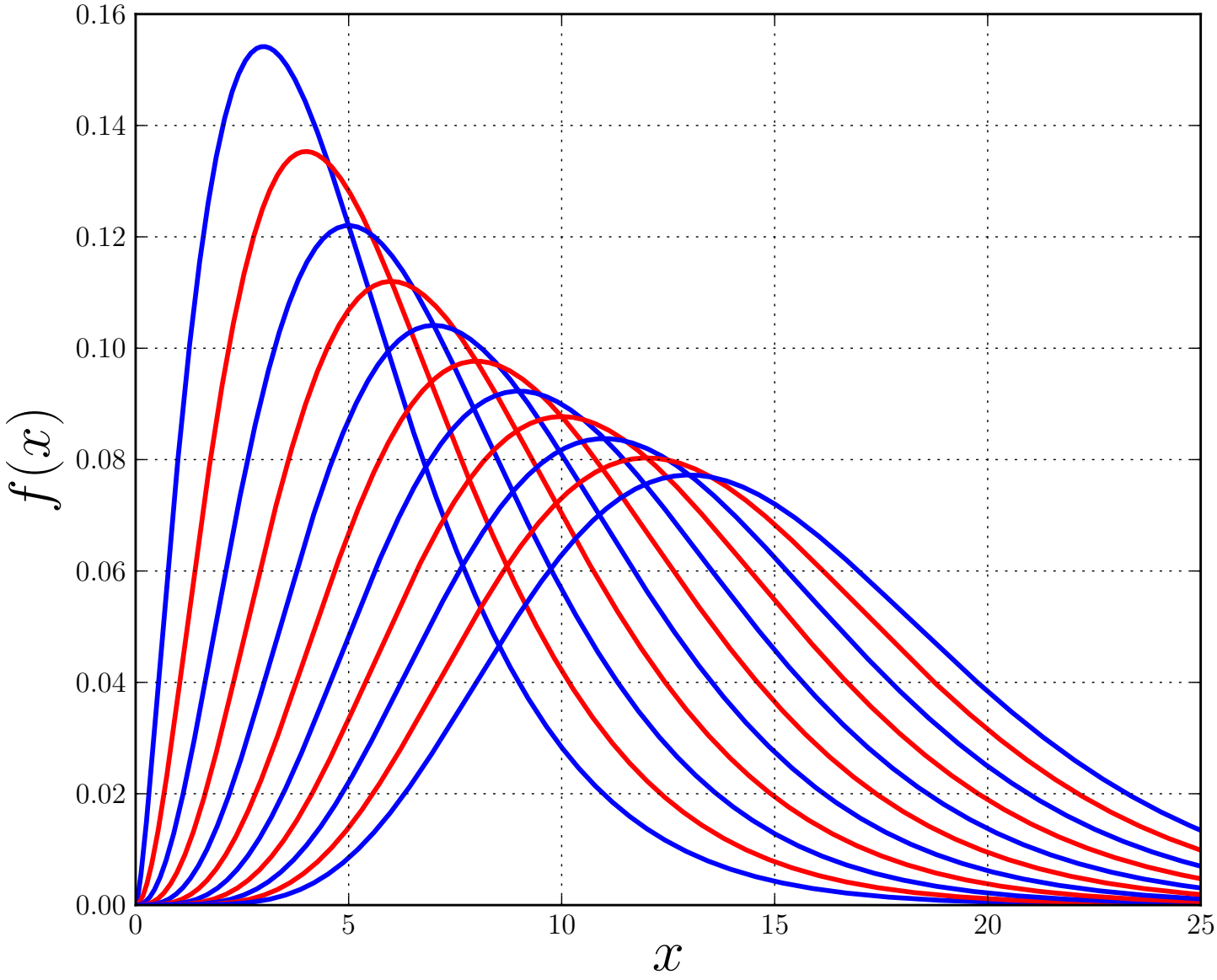
$$E[\chi_n^2] \;=\; E[X_1^2] \,+\, E[X_2^2] \,+\, \cdots \,+\, E[X_n^2] \;=\; n\,,$$

and since the $X_i$ are assumed *independent*,

$$Var[\chi_n^2] \;=\; Var[X_1^2] \,+\, Var[X_2^2] \,+\, \cdots \,+\, Var[X_n^2] \;=\; 2n\,,$$

and

$$\sigma(\chi_n^2) \;=\; \sqrt{2n}\,.$$




The Chi-Square *density* functions for $n = 5, 6, \cdots, 15$ .

( For *large* $n$ they look like *normal* density functions ! )

## The $\chi^2_n$ - Table

| $n$ | $\alpha = 0.975$ | $\alpha = \mathbf{0.95}$ | $\alpha = 0.05$ | $\alpha = 0.025$ |
|---|---|---|---|---|
| 5 | 0.83 | 1.15 | 11.07 | 12.83 |
| 6 | 1.24 | 1.64 | 12.59 | 14.45 |
| 7 | 1.69 | 2.17 | 14.07 | 16.01 |
| 8 | 2.18 | 2.73 | 15.51 | 17.54 |
| 9 | 2.70 | 3.33 | 16.92 | 19.02 |
| **10** | 3.25 | **3.94** | 18.31 | 20.48 |
| 11 | 3.82 | 4.58 | 19.68 | 21.92 |
| 12 | 4.40 | 5.23 | 21.03 | 23.34 |
| 13 | 5.01 | 5.89 | 22.36 | 24.74 |
| 14 | 5.63 | 6.57 | 23.69 | 26.12 |
| 15 | 6.26 | 7.26 | 25.00 | 27.49 |

This Table shows $z_{\alpha,n}$ values such that $P(\chi^2_n \geq z_{\alpha,n}) = \alpha$ .

( For example, $P(\chi^2_{\mathbf{10}} \geq \mathbf{3.94}) = \mathbf{95}\%$ )

# THE CENTRAL LIMIT THEOREM

The density function of the Chi-Square random variable

$$\chi^2_n \quad \equiv \quad \tilde{X}_1 \; + \; \tilde{X}_2 \; + \; \cdots \; + \; \tilde{X}_n \; ,$$

where

$$\tilde{X}_i = X_i^2 \; , \quad \text{and} \quad X_i \text{ is standard normal, } \; i = 1, 2, \cdots, n \; ,$$

starts looking like a *normal density function* when $n$ gets large.

- This remarkable fact holds much more generally !

- It is known as the *Central Limit Theorem* (CLT).

Let $X_1$ , $X_2$ , $\cdots$ , $X_n$ be *independent, identically distributed* ,

each having

$$\text{mean } \mu \quad , \quad \text{variance } \sigma^2 \quad , \quad \text{standard deviation } \sigma \ .$$

Then we know that

$$S \equiv X_1 + X_2 + \cdots + X_n \ ,$$

has

$$\textit{mean}: \qquad \mu_S \equiv E[S] = n\mu \qquad (\text{ Why ? })$$

$$\textit{variance}: \qquad Var(S) = n\sigma^2 \qquad (\text{ Why ? })$$

$$\textit{Standard deviation}: \qquad \sigma_S = \sqrt{n}\,\sigma$$

**NOTE** : $\sigma_S$ gets *bigger* as $n$ increases ( and so does $|\,\mu_S\,|$ ).

**THEOREM** (The Central Limit Theorem) (CLT) :

Let $X_1$ , $X_2$ , $\cdots$ , $X_n$ be *identical, independent* random variables,

each having

$$\text{mean } \mu \quad , \quad \text{variance } \sigma^2 \quad , \quad \text{standard deviation } \sigma \ .$$

Then for *large* $n$ the random variable

$$S \equiv X_1 + X_2 + \cdots + X_n \ ,$$

( which has mean $n\mu$ , variance $n\,\sigma^2$ , standard deviation $\sqrt{n}\,\sigma$ )

is *approximately normal* .

**NOTE** : Thus $\dfrac{S - n\mu}{\sqrt{n}\,\sigma}$ is approximately *standard normal* .

**EXAMPLE** : Recall that

$$\chi^2_n \quad \equiv \quad X_1^2 \; + \; X_2^2 \; + \; \cdots \; + \; X_n^2 \; ,$$

where each $X_i$ is standard normal, and (using moments) we found

$$\chi^2_n \text{ has } \textit{mean } n \quad \text{and} \quad \textit{standard deviation } \sqrt{2n} \; .$$

The Table below illustrates the accuracy of the approximation

$$P(\; \chi^2_n \leq 0 \;) \quad \cong \quad \Phi(\; \frac{0-n}{\sqrt{2n}} \;) \quad = \quad \Phi(-\sqrt{\frac{n}{2}} \;) \; .$$

| $n$ | $-\sqrt{\frac{n}{2}}$ | $\Phi(-\sqrt{\frac{n}{2}}\;)$ |
|---|---|---|
| 2 | $-1$ | 0.1587 |
| 8 | $-2$ | 0.0228 |
| 18 | $-3$ | 0.0013 |

**QUESTION** : What is the exact value of $P(\chi^2_n \leq 0)$ ? (!)

**EXERCISE** :

Use the approximation

$$P(\ \chi^2_n \leq x\ ) \quad \cong \quad \Phi(\ \frac{x-n}{\sqrt{2n}}\ )\ ,$$

to compute approximate values of

- $P(\ \chi^2_{32}\ \leq\ 24\ )$
- $P(\ \chi^2_{32}\ \geq\ 40\ )$
- $P(\ \ |\ \chi^2_{32} - 32\ |\ \ \leq\ 8\ )$

Let $X_1$ , $X_2$ , $\cdots$ , $X_n$ be *independent, identically distributed* ,

each having

$$\text{mean } \mu \text{ , variance } \sigma^2 \text{ , standard deviation } \sigma \text{ .}$$

Then we know that

$$\bar{X} \equiv \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right) ,$$

has

*mean* : $\mu_{\bar{X}} = E[\bar{X}] = \mu$ ( Why ? )

*variance* : $\sigma^2_{\bar{X}} = \frac{1}{n^2}\, n\sigma^2 = \frac{\sigma^2}{n}$ ( Why ? )

*Standard deviation* : $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

**NOTE** : $\sigma_{\bar{X}}$ gets *smaller* as $n$ increases.

**COROLLARY** (to the Central Limit Theorem) :

Let $X_1$ , $X_2$ , $\cdots$ , $X_n$ be *identical, independent* random variables,

each having

$$\text{mean} \quad \mu \quad , \quad \text{variance} \quad \sigma^2 \quad , \quad \text{standard deviation} \quad \sigma \ .$$

Then for *large* $n$ the random variable

$$\bar{X} \quad \equiv \quad \frac{1}{n} \left( X_1 \ + \ X_2 \ + \ \cdots \ + \ X_n \right) ,$$

( which has mean $\mu$ , variance $\dfrac{\sigma^2}{n}$ , standard deviation $\dfrac{\sigma}{\sqrt{n}}$ )

is *approximately normal* .

**NOTE** : Thus $\dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ is approximately *standard normal* .

**EXAMPLE** : Suppose $X_1$ , $X_2$ , $\cdots$ , $X_n$ are

*identical* , *independent* , *uniform* random variables ,

each having density function

$$f(x) \;=\; \frac{1}{2}\,, \quad \text{for} \quad x \in [-1,1]\,, \qquad (\ 0 \ \text{ otherwise }\ )\,,$$

with

*mean* $\mu = 0$ , *standard deviation* $\sigma = \dfrac{1}{\sqrt{3}}$ ( Check ! )

Then for *large* $n$ the random variable

$$\bar{X} \quad \equiv \quad \frac{1}{n}\,(X_1 \,+\, X_2 \,+\, \cdots \,+\, X_n)\,,$$

with

*mean* $\mu = 0$ , *standard deviation* $\sigma = \dfrac{1}{\sqrt{3n}}$ ,

is *approximately normal* , so that

$$P(\bar{X} \le x) \quad \cong \quad \Phi(\ \frac{x-0}{1/\sqrt{3n}}\ ) \quad \cong \quad \Phi(\ \sqrt{3n}\ x\ )\ .$$

**EXERCISE** : In the preceding example

$$P(\bar{X} \leq x) \quad \cong \quad \Phi\left( \frac{x - 0}{1/\sqrt{3n}} \right) \quad \equiv \quad \Phi( \sqrt{3n}\, x ) \ .$$

- Fill in the Table to illustrate the accuracy of this approximation :

| $n$ | $P(\bar{X} \leq -1) \ \cong \ \Phi(-\sqrt{3n}\ )$ |
|---|---|
| 3 | |
| 12 | |

( What is the exact value of $P(\ \bar{X} \leq -1\ )$ ? ! )

- For $n = 12$ find the approximate value of $P(\ \bar{X} \leq -0.1\ )$ .
- For $n = 12$ find the approximate value of $P(\ \bar{X} \leq -0.5\ )$ .

**EXPERIMENT** : ( a lengthy one $\cdots$ ! )

We give a detailed *computational example* to illustrate :

- The concept of *density function* .

- The numerical *construction* of a density function

  and (most importantly)

- The *Central Limit Theorem* .

## **EXPERIMENT** : ( continued $\cdots$ )

- Generate $N$ *uniformly distributed* random numbers in $[0, 1]$ .
- Many programming languages have a *function* for this.
- Call the random number values generated $\tilde{x}_i$ , $i = 1, 2, \cdots, N$ .
- Letting $x_i = 2\tilde{x}_i - 1$ gives *uniform random values in* $[-1, 1]$ .

## **EXPERIMENT** : ( continued $\cdots$ )

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| -0.737 | 0.511 | -0.083 | 0.066 | -0.562 | -0.906 | 0.358 | 0.359 |
| 0.869 | -0.233 | 0.039 | 0.662 | -0.931 | -0.893 | 0.059 | 0.342 |
| -0.985 | -0.233 | -0.866 | -0.165 | 0.374 | 0.178 | 0.861 | 0.692 |
| 0.054 | -0.816 | 0.308 | -0.168 | 0.402 | 0.821 | 0.524 | -0.475 |
| -0.905 | 0.472 | -0.344 | 0.265 | 0.513 | 0.982 | -0.269 | -0.506 |
| 0.965 | 0.445 | 0.507 | 0.303 | -0.855 | 0.263 | 0.769 | -0.455 |
| -0.127 | 0.533 | -0.045 | -0.524 | -0.450 | -0.281 | -0.667 | -0.027 |
| 0.795 | 0.818 | -0.879 | 0.809 | 0.009 | 0.033 | -0.362 | 0.973 |
| -0.012 | -0.468 | -0.819 | 0.896 | -0.853 | 0.001 | -0.232 | -0.446 |
| 0.828 | 0.059 | -0.071 | 0.882 | -0.900 | 0.523 | 0.540 | 0.656 |
| -0.749 | -0.968 | 0.377 | 0.736 | 0.259 | 0.472 | 0.451 | 0.999 |
| 0.777 | -0.534 | -0.387 | -0.298 | 0.027 | 0.182 | 0.692 | -0.176 |
| 0.683 | -0.461 | -0.169 | 0.075 | -0.064 | -0.426 | -0.643 | -0.693 |
| 0.143 | 0.605 | -0.934 | 0.069 | -0.003 | 0.911 | 0.497 | 0.109 |
| 0.781 | 0.250 | 0.684 | -0.680 | -0.574 | 0.429 | -0.739 | -0.818 |

120 values of a *uniform random variable* in $[-1, 1]$ .

**EXPERIMENT** : ( continued $\cdots$ )

- Divide $[-1,1]$ into $M$ subintervals of equal size $\Delta x = \frac{2}{M}$ .
- Let $I_k$ denote the $k$th interval, with midpoint $x_k$ .
- Let $m_k$ be the *frequency count* (# of random numbers in $I_k$) .
- Let $f(x_k) = \frac{m_k}{N\ \Delta x}$ , ($N$ is the total # of random numbers) .
- Then $\int_{-1}^{1} f(x)\,dx \cong \sum_{k=1}^{M} f(x_k)\,\Delta x = \sum_{k=1}^{M} \frac{m_k}{N\ \Delta x}\,\Delta x = 1$ ,
  and $f(x_k)$ approximates the value of the *density function* .
- The corresponding *distribution function* is

$$F(x_\ell) = \int_{-1}^{x_\ell} f(x)\ dx \cong \sum_{k=1}^{\ell} f(x_k)\ \Delta x\ .$$

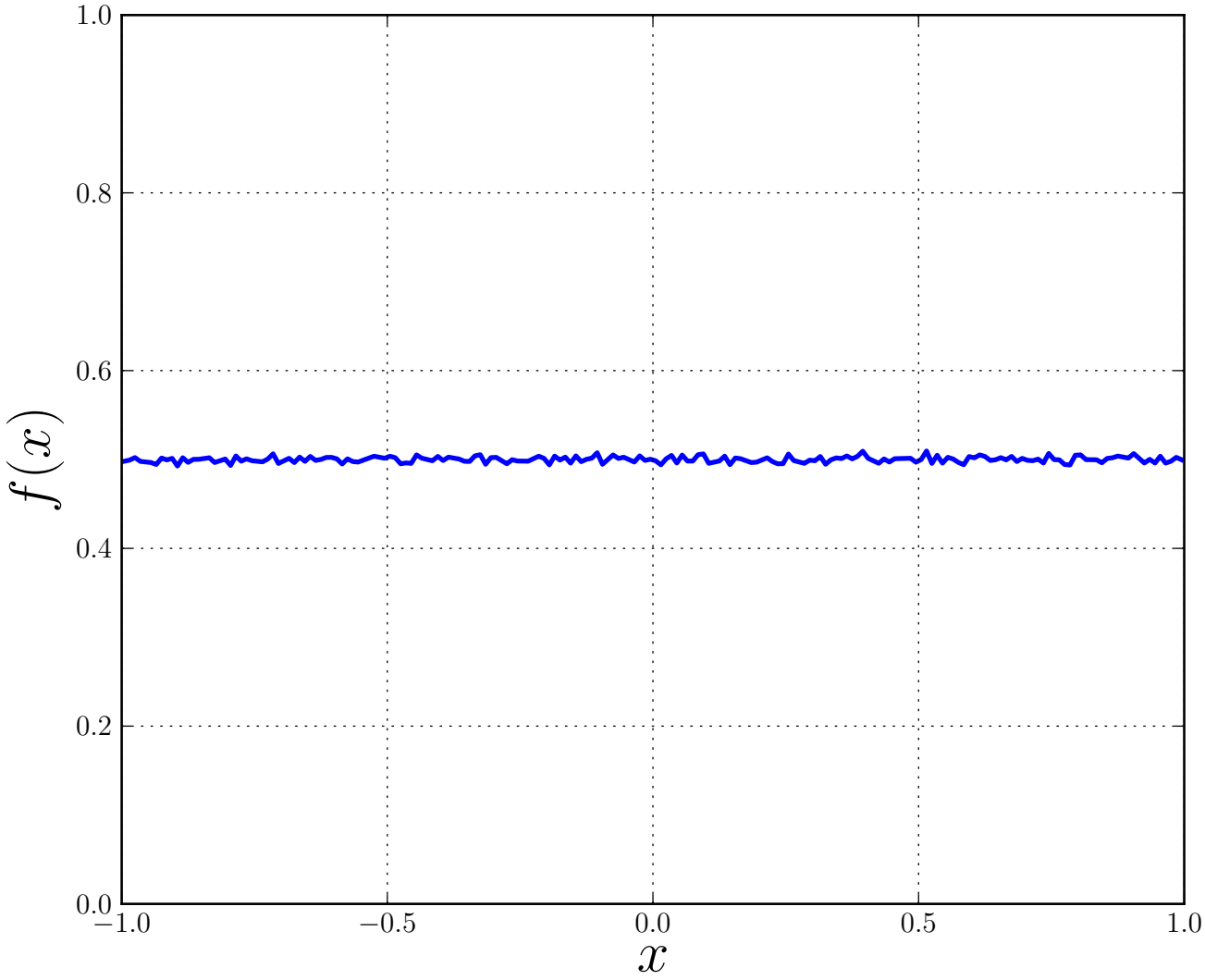
**EXPERIMENT** : ( continued $\cdots$ )

| Interval | Frequency | Sum | $f(x)$ | $F(x)$ |
|---|---|---|---|---|
| 1 | 50013 | 50013 | 0.500 | 0.067 |
| 2 | 50033 | 100046 | 0.500 | 0.133 |
| 3 | 50104 | 150150 | 0.501 | 0.200 |
| 4 | 49894 | 200044 | 0.499 | 0.267 |
| 5 | 50242 | 250286 | 0.502 | 0.334 |
| 6 | 49483 | 299769 | 0.495 | 0.400 |
| 7 | 50016 | 349785 | 0.500 | 0.466 |
| 8 | 50241 | 400026 | 0.502 | 0.533 |
| 9 | 50261 | 450287 | 0.503 | 0.600 |
| 10 | 49818 | 500105 | 0.498 | 0.667 |
| 11 | 49814 | 549919 | 0.498 | 0.733 |
| 12 | 50224 | 600143 | 0.502 | 0.800 |
| 13 | 49971 | 650114 | 0.500 | 0.867 |
| 14 | 49873 | 699987 | 0.499 | 0.933 |
| 15 | 50013 | 750000 | 0.500 | 1.000 |

*Frequency Table*, showing the count per interval .

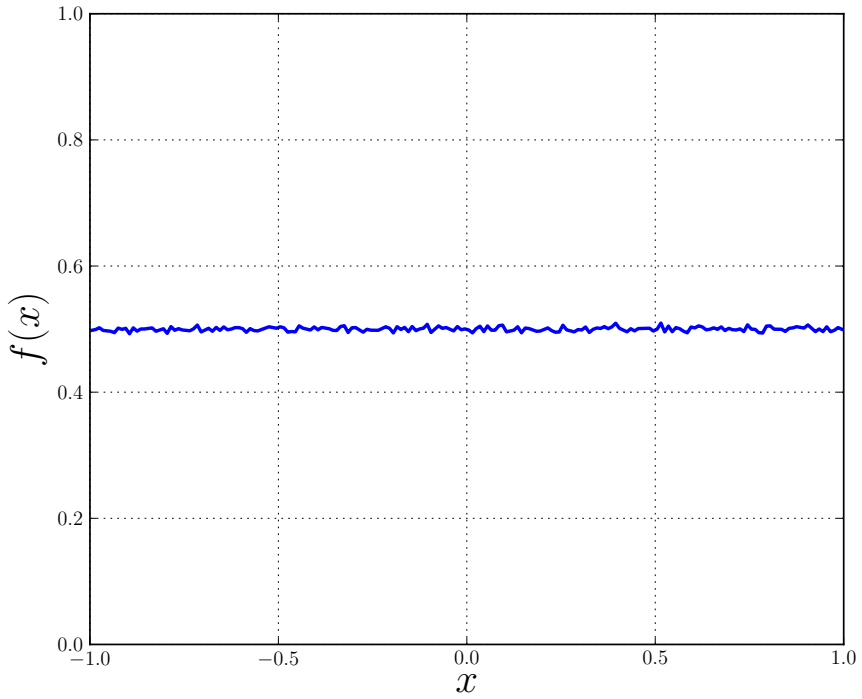
($N = 750,000$ random numbers, $M = 15$ intervals)

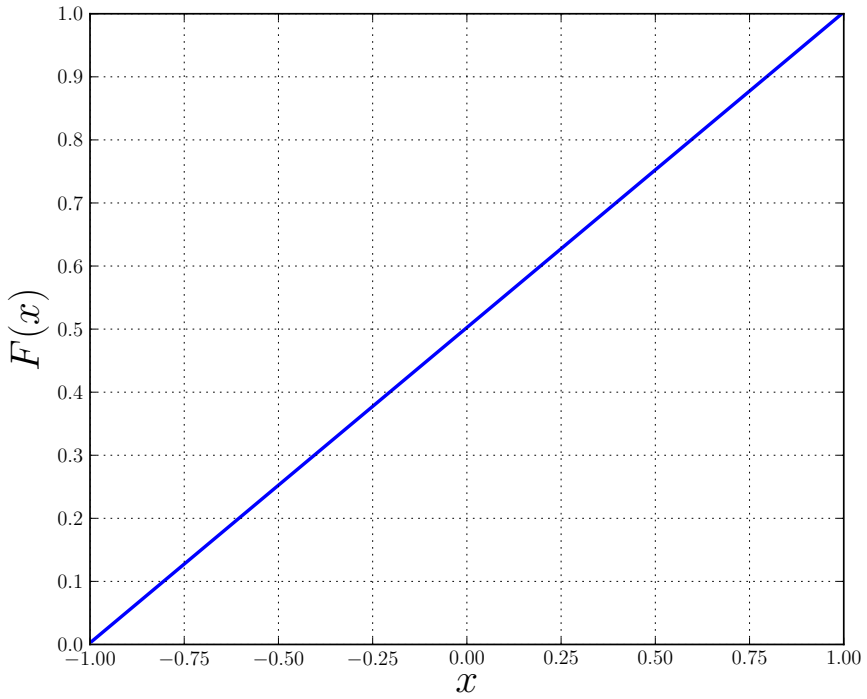
## EXPERIMENT : ( continued $\cdots$ )




The approximate *density function* , $f(x_k) = \frac{m_k}{N\ \Delta x}$ for $N = 5,000,000$ random numbers, and $M = 200$ intervals.

## EXPERIMENT : ( continued $\cdots$ )




Approximate *density function* $f(x)$ and *distribution function* $F(x)$ , for the case $N = 5,000,000$ random numbers, and $M = 200$ intervals.

**NOTE** : $F(x)$ appears *smoother* than $f(x)$ . ( Why ? )

**EXPERIMENT** : ( continued $\cdots$ )

Next $\cdots$ ( still for the uniform random variable in $[-1, 1]$ ) :

- Generate $n$ *random numbers* ( $n$ relatively *small* ) .
- Take the *average* of the $n$ random numbers.
- Do the above $N$ times, where (as before) $N$ is *very large* .
- Thus we deal with a random variable

$$\bar{X} \equiv \frac{1}{n} (X_1 + X_2 + \cdots + X_n) .$$

**EXPERIMENT** : ( continued $\cdots$ )

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| -0.047 | 0.126 | -0.037 | 0.148 | -0.130 | -0.004 | -0.174 | 0.191 |
| 0.198 | 0.073 | -0.025 | -0.070 | -0.018 | -0.031 | 0.063 | -0.064 |
| -0.197 | -0.026 | -0.062 | -0.004 | -0.083 | -0.031 | -0.102 | -0.033 |
| -0.164 | 0.265 | -0.274 | 0.188 | -0.067 | 0.049 | -0.090 | 0.002 |
| 0.118 | 0.088 | -0.071 | 0.067 | -0.134 | -0.100 | 0.132 | 0.242 |
| -0.005 | -0.011 | -0.018 | -0.048 | -0.153 | 0.016 | 0.086 | -0.179 |
| -0.011 | -0.058 | 0.198 | -0.002 | 0.138 | -0.044 | -0.094 | 0.078 |
| -0.011 | -0.093 | 0.117 | -0.156 | -0.246 | 0.071 | 0.166 | 0.142 |
| 0.103 | -0.045 | -0.131 | -0.100 | 0.072 | 0.034 | 0.176 | 0.108 |
| 0.108 | 0.141 | -0.009 | 0.140 | 0.025 | -0.149 | 0.121 | -0.120 |
| 0.012 | 0.002 | -0.015 | 0.106 | 0.030 | -0.096 | -0.024 | -0.111 |
| -0.147 | 0.004 | 0.084 | 0.047 | -0.048 | 0.018 | -0.183 | 0.069 |
| -0.236 | -0.217 | 0.061 | 0.092 | -0.003 | 0.005 | -0.054 | 0.025 |
| -0.110 | -0.094 | -0.115 | 0.052 | 0.135 | -0.076 | -0.018 | -0.121 |
| -0.030 | -0.146 | -0.155 | 0.089 | -0.177 | 0.027 | -0.025 | 0.020 |

Values of $\bar{X}$ for the case $N = 120$ and $n = 25$ .

**EXPERIMENT** : ( continued $\cdots$ )

For sample size $n$, $(n = 1, 2, 5, 10, 25)$ , and $M = 200$ intervals :

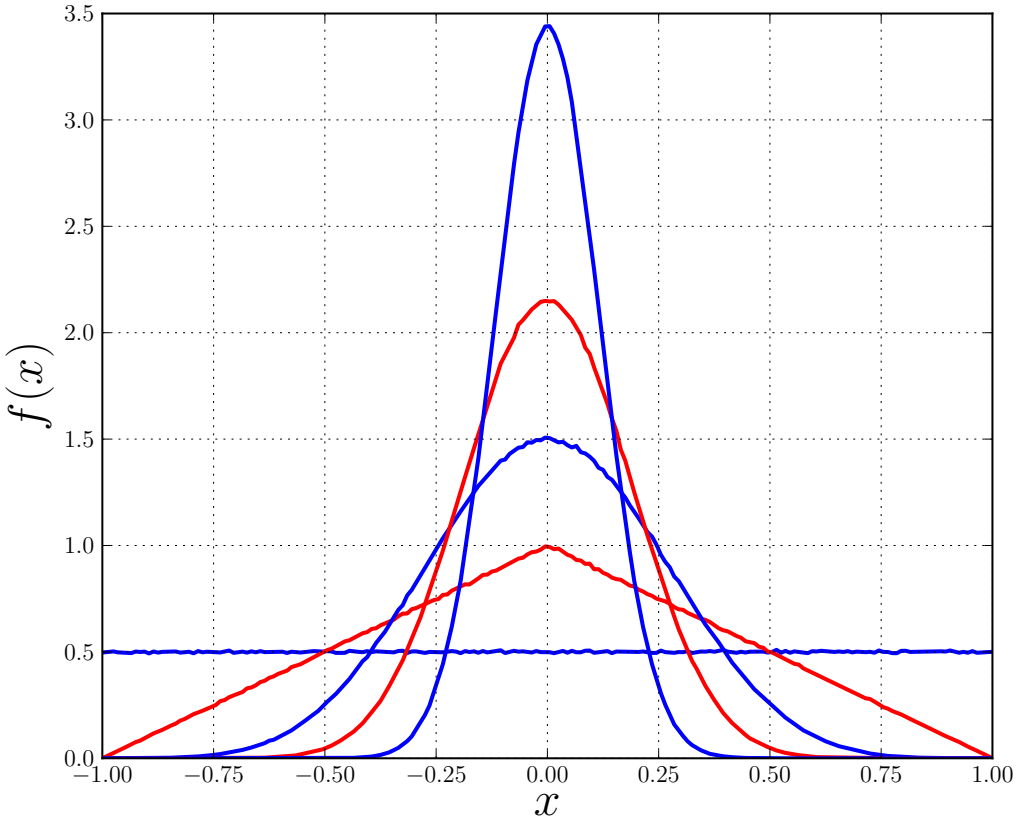
- Generate $N$ values of $\bar{X}$ , where $N$ *is very large* .
- Let $m_k$ be *the number of values of* $\bar{X}$ *in* $I_k$ .
- As before, let $f_n(x_k) = \frac{m_k}{N\ \Delta x}$ .
- Now $f_n(x_k)$ approximates the *density function of* $\bar{X}$ .

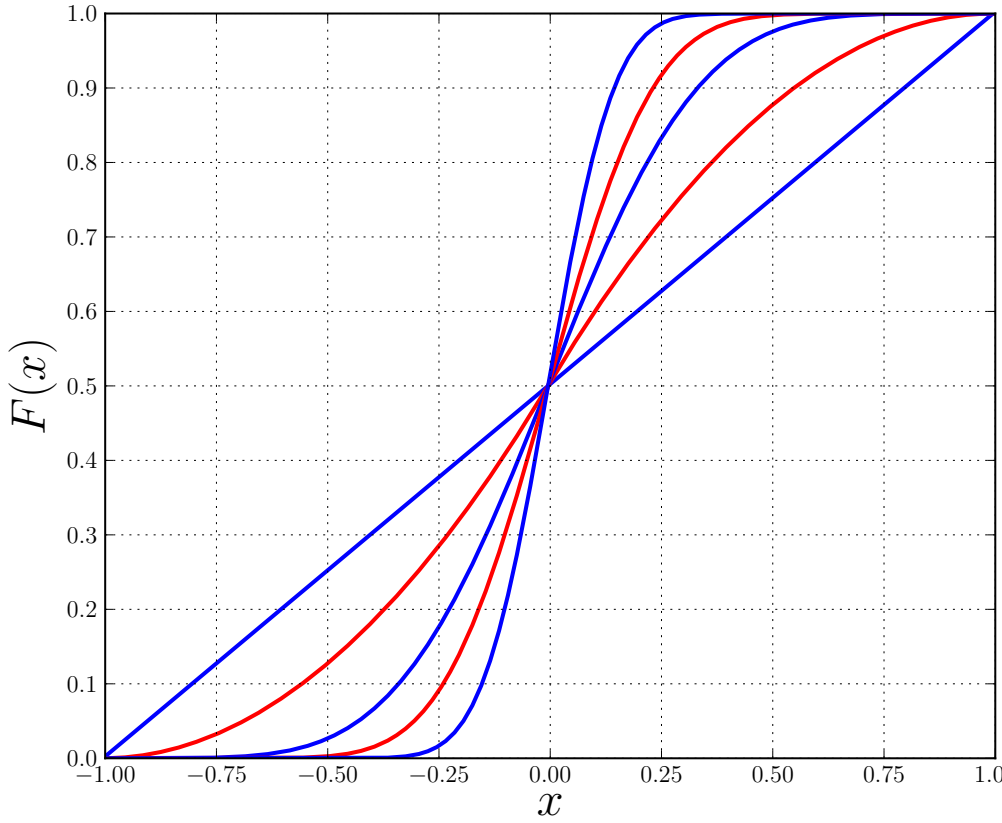
**EXPERIMENT** : ( continued $\cdots$ )

| Interval | Frequency | Sum | $f(x)$ | $F(x)$ |
|---|---|---|---|---|
| 1 | 0 | 0 | 0.00000 | 0.00000 |
| 2 | 0 | 0 | 0.00000 | 0.00000 |
| 3 | 0 | 0 | 0.00000 | 0.00000 |
| 4 | 11 | 11 | 0.00011 | 0.00001 |
| 5 | 1283 | 1294 | 0.01283 | 0.00173 |
| 6 | 29982 | 31276 | 0.29982 | 0.04170 |
| 7 | 181209 | 212485 | 1.81209 | 0.28331 |
| 8 | 325314 | 537799 | 3.25314 | 0.71707 |
| 9 | 181273 | 719072 | 1.81273 | 0.95876 |
| 10 | 29620 | 748692 | 0.29620 | 0.99826 |
| 11 | 1294 | 749986 | 0.01294 | 0.99998 |
| 12 | 14 | 750000 | 0.00014 | 1.00000 |
| 13 | 0 | 750000 | 0.00000 | 1.00000 |
| 14 | 0 | 750000 | 0.00000 | 1.00000 |
| 15 | 0 | 750000 | 0.00000 | 1.00000 |

*Frequency Table* for $\bar{X}$, showing the count per interval .
($N = 750,000$ values of $\bar{X}$, $M = 15$ intervals, sample size $n = 25$)

## EXPERIMENT : ( continued $\cdots$ )




The approximate *density functions* $f_n(x)$ , $n = 1,\ 2,\ 5,\ 10,\ 25$ ,

and the corresponding *distribution functions* $F_n(x)$ .

( $N = 5,000,000$ values of $\bar{X}$ , $M = 200$ intervals )

**EXPERIMENT** : ( continued $\cdots$ )

Recall that for *uniform random variables* $X_i$ on $[-1, 1]$

$$\bar{X} \quad \equiv \quad \frac{1}{n} \left( X_1 \, + \, X_2 \, + \, \cdots \, + \, X_n \right) ,$$

is *approximately normal* , with

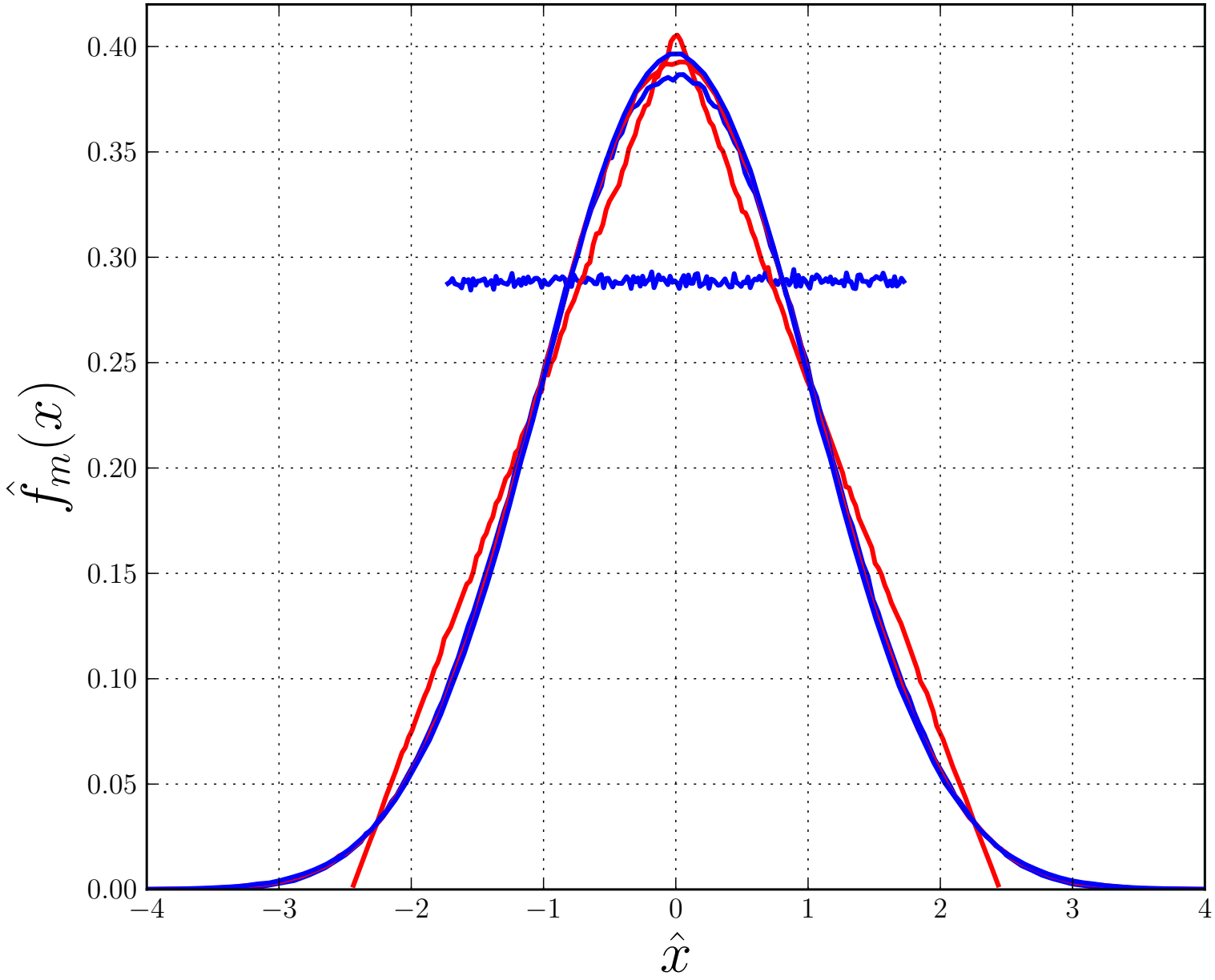
$$\textit{mean} \;\; \mu = 0 \quad , \quad \textit{standard deviation} \;\; \sigma = \frac{1}{\sqrt{3n}} \; .$$

Thus for each $n$ we can *normalize* $x$ and $f_n(x)$ :

$$\hat{x} \;\; = \;\; \frac{x - \mu}{\sigma} \;\; = \;\; \frac{x - 0}{\frac{1}{\sqrt{3n}}} \;\; = \;\; \sqrt{3n} \, x \qquad , \qquad \hat{f}_n(\hat{x}) \;\; = \;\; \frac{f_n(x)}{\sqrt{3n}} \; .$$

The next Figure shows :

- The normalized $\hat{f}_n(\hat{x})$ approach a limit as $n$ get large.
- This limit is the *standard normal* density function.
- Thus our computations agree with the Central Limit Theorem !




The normalized density functions $\hat{f}_n(x)$ , for $n = 1, 2, 5, 10, 25$ .

( $N = 5,000,000$ values of $\bar{X}$ , $M = 200$ intervals )

**EXERCISE** : Suppose

$$X_1 \ , \ X_2 \ , \ \cdots \ , \ X_{12} \ , \qquad (n = 12) \ ,$$

are identical, independent, *uniform* random variables on $[0, 1]$ .

We already know that each $X_i$ has

$$\text{mean} \ \ \mu = \tfrac{1}{2} \qquad , \qquad \text{standard deviation} \ \ \tfrac{1}{2\sqrt{3}} \ .$$

Let

$$\bar{X} \quad \equiv \quad \frac{1}{12} \left( X_1 \ + \ X_2 \ + \ \cdots \ + \ X_{12} \right) .$$

Use the CLT to *compute approximate values* of

- $P(\bar{X} \ \leq \ \frac{1}{3})$
- $P(\bar{X} \ \geq \ \frac{2}{3})$
- $P(|\ \bar{X} - \frac{1}{2}\ | \ \ \leq \ \frac{1}{3})$

**EXERCISE** : Suppose

$$X_1 \ , \ X_2 \ , \ \cdots \ , \ X_9 \ , \qquad (n = 9) \ ,$$

are identical, independent, *exponential* random variables, with

$$f(x) \ = \ \lambda e^{-\lambda x} \ , \quad \text{where} \quad \lambda = 1 \ .$$

We already know that each $X_i$ has

$$\text{mean} \ \ \mu = \tfrac{1}{\lambda} = 1 \ , \quad \text{and} \quad \text{standard deviation} \ \ \tfrac{1}{\lambda} = 1 \ .$$

Let

$$\bar{X} \ \equiv \ \frac{1}{9} \left( X_1 \ + \ X_2 \ + \ \cdots \ + \ X_9 \right) .$$

Use the CLT to *compute approximate values* of

- $P(\bar{X} \ \leq \ 0.4)$
- $P(\bar{X} \ \geq \ 1.6)$
- $P(\mid \bar{X} - 1 \mid \ \leq \ 0.6)$

**EXERCISE** : Suppose

$$X_1 \ , \ X_2 \ , \ \cdots \ , \ X_n \ ,$$

are identical, independent, *normal* random variables, with

$$\text{mean} \ \ \mu = 7 \qquad , \qquad \text{standard deviation} \ \ 4 \ .$$

Let

$$\bar{X} \quad \equiv \quad \frac{1}{n} \left( X_1 \ + \ X_2 \ + \ \cdots \ + \ X_n \right) .$$

Use the CLT to determine at least how big $n$ must be so that

- $P( \mid \bar{X} - \mu \mid \ \leq \ 1) \ \ \geq \ \ 90 \ \% \ .$

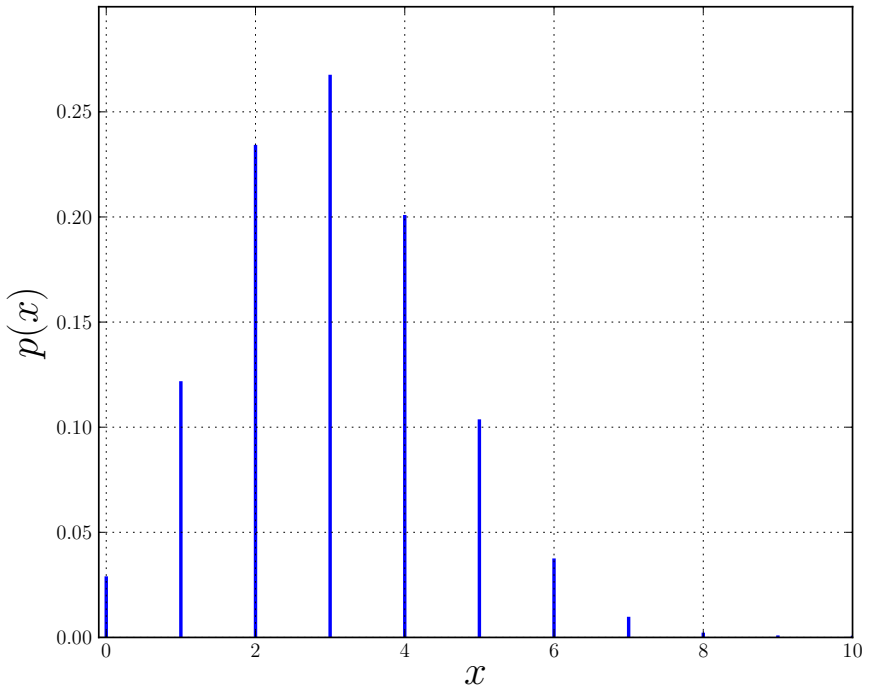
**EXAMPLE** : The CLT also applies to *discrete random variables* .

The *Binomial random variable* , with

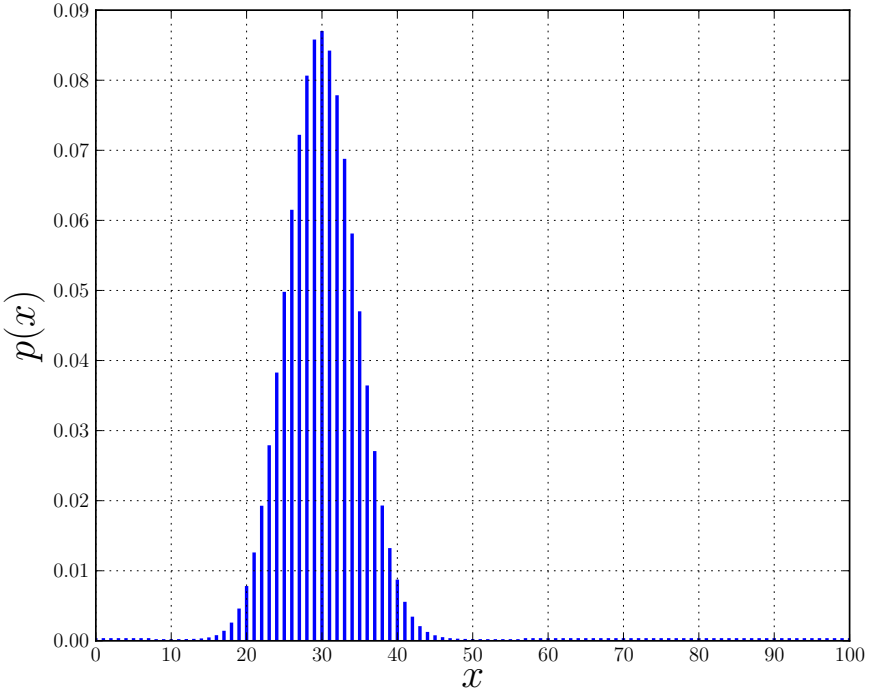
$$P(X = k) \quad = \quad \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \ , \qquad (0 \leq k \leq n) \ ,$$

is already a *sum* (namely, of *Bernoulli* random variables).

Thus its *binomial probability mass function* already "*looks normal*" :




Binomial : $n = 10$ , $p = 0.3$




Binomial : $n = 100$ , $p = 0.3$

**EXAMPLE** : ( continued $\cdots$ )

We already know that if $X$ is *binomial* then

$$\mu(X) \;=\; np \qquad \text{and} \qquad \sigma(X) \;=\; \sqrt{np(1-p)}\;.$$

Thus, for $n = 100$ , $p = 0.3$ , we have

$$\mu(X) \;=\; 30 \qquad \text{and} \qquad \sigma(X) \;=\; \sqrt{21} \;\cong\; 4.58\;.$$

Using the CLT we can *approximate*

$$P(X \leq 26) \;\cong\; \Phi(\frac{26-30}{4.58}) \;=\; \Phi(-0.87) \;\cong\; \mathbf{19.2}\;\%\;.$$

The *exact binomial value* is

$$P(X \leq 26) \;=\; \sum_{k=0}^{26} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \;=\; \mathbf{22.4}\;\%\;,$$

**QUESTION** : What do you say ?

**EXAMPLE** : ( continued $\cdots$ )

We found the *exact* binomial value

$$P(X \leq 26) \quad = \quad \mathbf{22.4}\ \% \ ,$$

and the CLT *approximation*

$$P(X \leq 26) \quad \cong \quad \Phi(\frac{26-30}{4.58}) \quad = \quad \Phi(-0.87) \quad \cong \quad 19.2\ \% \ .$$

It is *better* to

"*spread*" $P(X = 26)$ *over the interval* $[25.5\ ,\ 26.5]$ . ( Why ? )

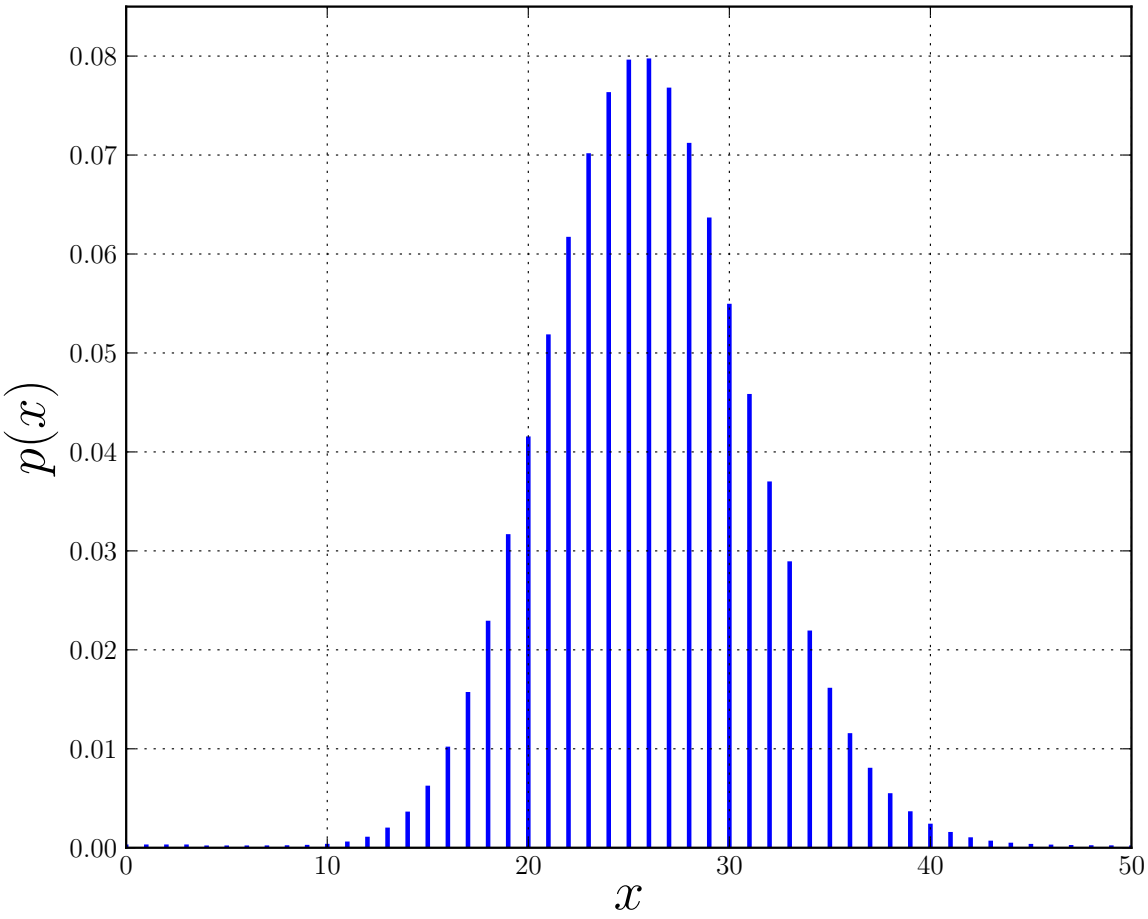
Thus it is *better* to *adjust* the approximation to $P(X \leq 26)$ by

$$P(X \leq 26) \quad \cong \quad \Phi(\frac{26.5-30}{4.58}) \quad = \quad \Phi(-0.764) \quad \cong \quad \mathbf{22.2}\ \% \ .$$

**QUESTION** : What do you say now ?

**EXERCISE** :

Consider the *Binomial* distribution with $n = 676$ and $p = \frac{1}{26}$ :




The Binomial $(n = 676,\ p = \frac{1}{26})$, shown in $[0, 50]$ .

**EXERCISE** : (continued $\cdots$) (Binomial : $n = 676$ , $p = \frac{1}{26}$ )

- Write down the *Binomial formula* for $P(X = 24)$ .
- Evaluate $P(X = 24)$ using the Binomial *recurrence formula* .
- Compute $E[X] = np$ and $\sigma(X) = \sqrt{np(1-p)}$ .

The *Poisson* probability mass function

$$P(X = k) \;=\; e^{-\lambda}\,\frac{\lambda^k}{k!}\;, \qquad (\text{with } \lambda = np\;)\;,$$

approximates the Binomial when $p$ is small and $n$ large.

- Evaluate $P(X = 24)$ using the Poisson *recurrence formula* .
- Compute the *standard normal* approximation to $P(X = 24)$.

**ANSWERS** : 7.61 % , 7.50 % , 7.36 % .

**EXPERIMENT** :

Compare the *accuracy* of the Poisson and the adjusted Normal *approximations* to the Binomial, for different values of $n$ .

| $k$ | $n$ | Binomial | Poisson | Normal |
|---|---|---|---|---|
| 2 | 4 | 0.6875 | 0.6767 | 0.6915 |
| 4 | 8 | 0.6367 | 0.6288 | 0.6382 |
| 8 | 16 | 0.5982 | 0.5925 | 0.5987 |
| 16 | 32 | 0.5700 | 0.5660 | 0.5702 |
| 32 | 64 | 0.5497 | 0.5468 | 0.5497 |
| 64 | 128 | 0.5352 | 0.5332 | 0.5352 |

$P(X \leq k)$ , where $k = \lfloor np \rfloor$, with $p = 0.5$ .

- Any conclusions ?

## EXPERIMENT : ( continued $\cdots$ )

Compare the *accuracy* of the Poisson and the adjusted Normal *approximations* to the Binomial, for different values of $n$ .

| $k$ | $n$ | Binomial | Poisson | Normal |
|---|---|---|---|---|
| 0 | 4 | 0.6561 | 0.6703 | 0.5662 |
| 0 | 8 | 0.4305 | 0.4493 | 0.3618 |
| 1 | 16 | 0.5147 | 0.5249 | 0.4668 |
| 3 | 32 | 0.6003 | 0.6025 | 0.5702 |
| 6 | 64 | 0.5390 | 0.5423 | 0.5166 |
| 12 | 128 | 0.4805 | 0.4853 | 0.4648 |
| 25 | 256 | 0.5028 | 0.5053 | 0.4917 |
| 51 | 512 | 0.5254 | 0.5260 | 0.5176 |

$P(X \leq k)$ , where $k = \lfloor np \rfloor$, with $p = 0.1$ .

- Any conclusions ?

## EXPERIMENT : ( continued $\cdots$ )

Compare the *accuracy* of the Poisson and the adjusted Normal *approximations* to the Binomial, for different values of $n$ .

| $k$ | $n$ | Binomial | Poisson | Normal |
|---|---|---|---|---|
| 0 | 4 | 0.9606 | 0.9608 | 0.9896 |
| 0 | 8 | 0.9227 | 0.9231 | 0.9322 |
| 0 | 16 | 0.8515 | 0.8521 | 0.8035 |
| 0 | 32 | 0.7250 | 0.7261 | 0.6254 |
| 0 | 64 | 0.5256 | 0.5273 | 0.4302 |
| 1 | 128 | 0.6334 | 0.6339 | 0.5775 |
| 2 | 256 | 0.5278 | 0.5285 | 0.4850 |
| 5 | 512 | 0.5948 | 0.5949 | 0.5670 |
| 10 | 1024 | 0.5529 | 0.5530 | 0.5325 |
| 20 | 2048 | 0.5163 | 0.5165 | 0.5018 |
| 40 | 4096 | 0.4814 | 0.4817 | 0.4712 |

$P(X \leq k)$ , where $k = \lfloor np \rfloor$, with $p = 0.01$ .

- Any conclusions ?

# SAMPLE STATISTICS

*Sampling* can consist of

- *Gathering random data* from a large *population*, for example,
  - measuring the height of randomly selected adults
  - measuring the starting salary of random CS graduates

- Recording the *results of experiments* , for example,
  - measuring the breaking strength of randomly selected bolts
  - measuring the lifetime of randomly selected light bulbs

- We shall generally assume the population is *infinite* (or *large*) .
- We shall also generally assume the observations are *independent* .
- The outcome of any experiment does not affect other experiments.

**DEFINITIONS** :

- A *random sample* from a population consists of

  *independent* , *identically distributed* random variables,

  $$X_1 \ , \ X_2 \ , \ \cdots \ , \ X_n \ .$$

- The values of the $X_i$ are called the *outcomes* of the experiment.

- A *statistic* is a *function* of $X_1, \ X_2, \ \cdots \ , X_n$ .

- Thus a *statistic* itself is a *random variable* .

**EXAMPLES** :

The most *important statistics* are

- The *sample mean*

$$\bar{X} \quad \equiv \quad \frac{1}{n}\,(X_1 \,+\, X_2 \,+\, \cdots \,+\, X_n)\;.$$

- The *sample variance*

$$S^2 \quad \equiv \quad \frac{1}{n}\,\sum_{k=1}^{n}(X_k - \bar{X})^2\;.$$

(to be discussed in detail $\cdots$ )

- The *sample standard deviation* $S \;=\; \sqrt{S^2}\;.$

For a random sample

$$X_1 \ , \ X_2 \ , \ \cdots \ , \ X_n \ ,$$

one can think of many other *statistics* such as :

- The *order statistic* in which the observation are *ordered in size* .

- The *sample median* , which is
  - the *midvalue* of the order statistic (if $n$ is odd),
  - the *average* of the *two middle values* (if $n$ is even).

- The *sample range* : the difference between the largest and smallest observation.

**EXAMPLE** : For the 8 observations

$$-0.737\ ,\ 0.511\ ,-0.083\ ,\ 0.066\ ,-0.562\ ,-0.906\ ,\ 0.358\ ,\ 0.359\ ,$$

from the first row of the Table given earlier, we have

*Sample mean* :

$$\bar{X} \ = \ \frac{1}{8}\,(\ -0.737\ +\ 0.511\ -\ 0.083\ +\ 0.066$$
$$-\ 0.562\ -\ 0.906\ +\ 0.358\ +\ 0.359\ )\ \ =\ \ -0.124\ .$$

*Sample variance* :

$$\frac{1}{8}\{\ (-0.737-\bar{X})^2\ +\ (0.511-\bar{X})^2\ +\ (-0.083-\bar{X})^2$$
$$+\ (0.066-\bar{X})^2\ +\ (-0.562-\bar{X})^2\ +\ (-0.906-\bar{X})^2$$
$$+\ (0.358-\bar{X})^2\ +\ (0.359-\bar{X})^2\ \}\ \ =\ \ 0.26\ .$$

*Sample standard deviation* : $\sqrt{0.26}\ \ =\ \ 0.51\ .$

**EXAMPLE** : ( continued $\cdots$ )

For the 8 observations

$$-0.737\ ,\ 0.511\ , -0.083\ ,\ 0.066\ , -0.562\ , -0.906\ ,\ 0.358\ ,\ 0.359\ ,$$

we also have

The *order statistic* :

$$-0.906\ , -0.737\ , -0.562\ , -0.083\ ,\ 0.066\ ,\ 0.358\ ,\ 0.359\ ,\ 0.511\ .$$

The *sample median* : $(-0.083 + 0.066)/2 \ = \ -\,0.0085\ .$

The *sample range* : $0.511 - (-0.906) \ = \ 1.417\ .$

## The Sample Mean

Suppose the *population mean* and *standard deviation* are $\mu$ and $\sigma$ .

The *sample mean*

$$\bar{X} \equiv \frac{1}{n}\,(X_1 + X_2 + \cdots + X_n)\,,$$

is also a *random variable* , and, as before, we have

$$\mu_{\bar{X}} \equiv E[\bar{X}] = E[\,\frac{1}{n}\,(X_1 + X_2 + \cdots + X_n)\,] = \mu\,,$$

$$\sigma^2_{\bar{X}} \equiv Var(\bar{X}) \equiv E[(\bar{X} - \mu)^2] = \frac{\sigma^2}{n}\,,$$

*Standard deviation* $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$ .

**NOTE** : The *sample mean* approximates the *population mean* $\mu$ .

How well does the *sample mean* approximate the *population mean* ?

From the Corollary to the CLT we know

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

is approximately *standard normal* when $n$ is large.

Thus, for given $n$ and $z$ , $(z > 0)$ , we can, for example, estimate

$$P(\ |\ \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\ |\ \leq\ z\ )\ \cong\ 1\ -\ 2\ \Phi(-z)\ .$$

(A problem is that we often don't know the value of $\sigma$ $\cdots$ )

Now

$$
\begin{aligned}
|\,\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\,| \;\le\; z \quad &\Longleftrightarrow \quad -z \;\le\; \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \;\le\; z \\
&\Longleftrightarrow \quad -\frac{\sigma z}{\sqrt{n}} \;\le\; \bar{X}-\mu \;\le\; \frac{\sigma z}{\sqrt{n}} \\
&\Longleftrightarrow \quad -\bar{X}-\frac{\sigma z}{\sqrt{n}} \;\le\; -\mu \;\le\; -\bar{X}+\frac{\sigma z}{\sqrt{n}} \\
&\Longleftrightarrow \quad \bar{X}-\frac{\sigma z}{\sqrt{n}} \;\le\; \mu \;\le\; \bar{X}+\frac{\sigma z}{\sqrt{n}} \qquad (\text{ Why ? })
\end{aligned}
$$

Thus

$$
P(\,\mu \in [\bar{X}-\frac{\sigma z}{\sqrt{n}}\,,\,\bar{X}+\frac{\sigma z}{\sqrt{n}}]\,) \;=\; P(\;|\,\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\,|\;\le\; z\,) \;\cong\; 1-2\,\Phi(-z)\,.
$$

We found : $$P(\ \mu \in [\bar{X} - \frac{\sigma z}{\sqrt{n}}\ ,\ \bar{X} + \frac{\sigma z}{\sqrt{n}}]\ ) \quad \cong \quad 1 - 2\ \Phi(-z)\ .$$

**EXAMPLE** : We take samples from a given population :

- The population *mean* $\mu$ is *unknown* .
- The population standard deviation is $\sigma = 3$
- The sample size is $n = 25$ .
- The sample mean is $\bar{X} = 4.5$ .

Taking **z=2** , we have

$$P(\ \mu \in [\ 4.5 - \frac{3 \cdot \mathbf{2}}{\sqrt{25}}\ ,\ 4.5 + \frac{3 \cdot \mathbf{2}}{\sqrt{25}}\ ]\ ) \quad = \quad P(\ \mu \in [\ 3.3\ ,\ 5.7\ ]\ )$$

$$\cong \quad 1 - 2\ \Phi(\mathbf{-2}) \quad \cong \quad 95\ \%\ .$$

We call [ 3.3 , 5.7 ] the 95 % *confidence interval estimate* of $\mu$ .

**EXERCISE** :

As in the preceding example, $\mu$ is unknown, $\sigma = 3$ , $\bar{X} = 4.5$ .

Use the formula

$$P(\ \mu \in [\bar{X} - \frac{\sigma z}{\sqrt{n}}\ ,\ \bar{X} + \frac{\sigma z}{\sqrt{n}}]\ ) \quad \cong \quad 1 - 2\Phi(-z)\ ,$$

to determine

- The 50 % *confidence interval estimate* of $\mu$ when $n = 25$ .
- The 50 % *confidence interval estimate* of $\mu$ when $n = 100$ .
- The 95 % *confidence interval estimate* of $\mu$ when $n = 100$ .

**NOTE** : In the *Standard Normal Table* , check that

- The 50 % confidence interval corresponds to $z = 0.68 \cong 0.7$ .
- The 95 % confidence interval corresponds to $z = 1.96 \cong 2.0$ .

**The Sample Variance** We defined the *sample variance* as

$$S^2 \equiv \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2 = \sum_{k=1}^{n}\left[\,(X_k - \bar{X})^2 \cdot \frac{1}{n}\,\right].$$

Earlier, for *discrete* random variables $X$, we defined the *variance* as

$$\sigma^2 \equiv E[\,(X-\mu)^2] \equiv \sum_{k}\left[\,(X_k - \mu)^2 \cdot p(X_k)\,\right].$$

- These two formulas look *deceptively* similar !
- In fact, they are quite different !
- The 1st sum for $S^2$ is *only* over the *sampled* $X$-values.
- The 2nd sum for $\sigma^2$ is over *all* $X$-values.
- The 1st sum for $S^2$ has *constant weights* .
- The 2nd sum for $\sigma^2$ uses the *probabilities as weights* .

We have just argued that the *sample variance*

$$S^2 \equiv \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2 ,$$

and the *population variance* (for *discrete* random variables)

$$\sigma^2 \equiv E[\,(X-\mu)^2] \equiv \sum_k [\,(X_k-\mu)^2 \cdot p(X_k)\,] ,$$

are quite different.

Nevertheless, we will show that *for large $n$ their values are close* !

Thus for large $n$ we have the approximation

$$S^2 \cong \sigma^2 .$$

**FACT 1** : We (obviously) have that

$$\bar{X} \;=\; \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \text{implies} \qquad \sum_{k=1}^{n} X_k \;=\; n\bar{X} \;.$$

**FACT 2** : From

$$\sigma^2 \;\equiv\; Var(X) \;\equiv\; E[(X-\mu)^2] \;=\; E[X^2] \;-\; \mu^2 \;,$$

we (obviously) have

$$E[X^2] \;=\; \sigma^2 \;+\; \mu^2 \;.$$

**FACT 3** : Recall that for *independent, identically distributed* $X_k$ ,

where each $X_k$ has *mean* $\mu$ and *variance* $\sigma^2$ , we have

$$\mu_{\bar{X}} \;\equiv\; E[\bar{X}] \;=\; \mu \qquad , \qquad \sigma^2_{\bar{X}} \;\equiv\; E[(\bar{X}-\mu)^2] \;=\; \frac{\sigma^2}{n} \;.$$

**FACT 4** :

$$S^2 \equiv \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2 = \frac{1}{n}\left[\sum_{k=1}^{n} X_k^2\right] - \bar{X}^2 .$$

**PROOF** :

$$S^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

$$= \frac{1}{n}\sum_{k=1}^{n}(X_k^2 - 2X_k\bar{X} + \bar{X}^2)$$

$$= \frac{1}{n}\left[\sum_{k=1}^{n} X_k^2 - 2\bar{X}\sum_{k=1}^{n} X_k + n\bar{X}^2\right] \qquad \text{(now use Fact 1)}$$

$$= \frac{1}{n}\left[\sum_{k=1}^{n} X_k^2 - 2n\bar{X}^2 + n\bar{X}^2\right] = \frac{1}{n}\left[\sum_{k=1}^{n} X_k^2\right] - \bar{X}^2 \quad \textbf{QED} !$$

**THEOREM** : The sample variance

$$S^2 \equiv \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

has *expected value*

$$E[S^2] = (1 - \frac{1}{n}) \cdot \sigma^2 \ .$$

**PROOF** :

$$E[S^2] = E[\ \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2\ ]$$

$$= E\Big[\ \frac{1}{n}\sum_{k=1}^{n}[X_k^2\ ] - \bar{X}^2\ \Big] \qquad \text{(using Fact 4)}$$

$$= \frac{1}{n}\sum_{k=1}^{n} E[X_k^2] - E[\bar{X}^2]$$

$$= \sigma^2 + \mu^2 - (\sigma_{\bar{X}}^2 + \mu_{\bar{X}}^2) \qquad \text{(using Fact 2} \quad n+1 \text{ times !)}$$

$$= \sigma^2 + \mu^2 - (\frac{\sigma^2}{n} + \mu^2) = (1 - \frac{1}{n})\,\sigma^2 \ . \quad \text{(Fact 3)} \qquad \textbf{QED} \ !$$

**REMARK** : Thus $\lim_{n\to\infty} E[S^2] = \sigma^2$ .

Most authors instead define the *sample variance* as

$$\hat{S}^2 \equiv \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 \ .$$

In this case the Theorem becomes :

**THEOREM** : The sample variance

$$\hat{S}^2 \equiv \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

has *expected value*

$$E[\hat{S}^2] = \sigma^2 \ .$$

**EXERCISE** : Check this !

**EXAMPLE** : The *random sample* of 120 values of a *uniform random variable* on $[-1, 1]$ in an earlier Table has

$$\begin{aligned} \bar{X} &= \frac{1}{120} \sum_{k=1}^{120} X_k = 0.030 \ , \\ S^2 &= \frac{1}{120} \sum_{k=1}^{120} (X_k - \bar{X})^2 = 0.335 \ , \\ S &= \sqrt{S^2} = 0.579 \ , \end{aligned}$$

while

$$\begin{aligned} \mu &= 0 \ , \\ \sigma^2 &= \int_{-1}^{1} (x - \mu)^2 \, \frac{1}{2} \, dx = \frac{1}{3} \ , \\ \sigma &= \sqrt{\sigma^2} = \frac{1}{\sqrt{3}} = 0.577 \ . \end{aligned}$$

- What do you say ?

**EXAMPLE** :

- Generate 50 *uniform random numbers* in $[-1, 1]$ .
- Compute their average.
- Do the above 500 times.
- Call the results $\bar{X}_k$ , $k = 1, 2, \cdots, 500$ .
- Thus each $\bar{X}_k$ is the *average* of 50 random numbers.
- Compute the *sample statistics* $\bar{X}$ and $S$ of these 500 values.
- Can you *predict* the values of $\bar{X}$ and $S$ ?

**EXAMPLE** : ( continued $\cdots$ )

Results : $$\bar{X} \;=\; \frac{1}{500}\sum_{k=1}^{500}\bar{X}_k \;=\; -\,0.00136\ ,$$

$$S^2 \;=\; \frac{1}{500}\sum_{k=1}^{500}(\bar{X}_k-\bar{X})^2 \;=\; 0.00664\ ,$$

$$S \;=\; \sqrt{S^2} \;=\; 0.08152\ .$$

**EXERCISE** :

- What is the value of $E[\bar{X}]$ ?
- Compare $\bar{X}$ to $E[\bar{X}]$ .
- What is the value of $Var(\bar{X})$ ?
- Compare $S^2$ to $Var(\bar{X})$ .

## Estimating the variance of a normal distribution

We have shown that

$$S^2 \equiv \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2 \cong \sigma^2 .$$

How good is this approximation for *normal random variables* $X_k$ ?

To answer this we first need :

**FACT 1** :

$$\sum_{k=1}^{n}(X_k - \mu)^2 - \sum_{k=1}^{n}(X_k - \bar{X})^2 = n(\bar{X} - \mu)^2 .$$

**PROOF** :

$$\begin{aligned}
\text{LHS} &= \textstyle\sum_{k=1}^{n}\{\, X_k^2 - 2X_k\mu + \mu^2 - X_k^2 + 2X_k\bar{X} - \bar{X}^2 \,\} \\
&= -2n\bar{X}\mu + n\mu^2 + 2n\bar{X}^2 - n\bar{X}^2 \\
&= n\bar{X}^2 - 2n\bar{X}\mu + n\mu^2 = \text{RHS} .
\end{aligned}$$

**QED** !

Rewrite Fact 1

$$\sum_{k=1}^{n}(X_k-\mu)^2 \quad - \quad \sum_{k=1}^{n}(X_k-\bar{X})^2 \quad = \; n(\bar{X}-\mu)^2 \; ,$$

as

$$\sum_{k=1}^{n}\Big(\frac{X_k-\mu}{\sigma}\Big)^2 \quad - \quad \frac{n}{\sigma^2}\,\frac{1}{n}\,\sum_{k=1}^{n}(X_k-\bar{X})^2 \; = \; \Big(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\Big)^2 \; ,$$

and then as

$$\sum_{k=1}^{n} Z_k^2 \quad - \quad \frac{n}{\sigma^2}\,S^2 \quad = \quad Z^2 \; ,$$

where

$$S^2 \;\; \text{is the sample variance} \; ,$$

and

$Z$ and $Z_k$ are *standard normal* because the $X_k$ are *normal* .

Finally, we can write the above as

$$\frac{n}{\sigma^2}\,S^2 \quad = \quad \chi_n^2 \; - \; \chi_1^2 \; . \qquad (\text{ Why ? })$$

We have found that

$$\frac{n}{\sigma^2}\, S^2 \quad = \quad \chi^2_n \;-\; \chi^2_1 \;.$$

**THEOREM** : For samples from a normal distribution :

$$\frac{n}{\sigma^2}\, S^2 \;\text{ has the }\; \chi^2_{n-1} \;\text{ distribution !}$$

**PROOF** : Omitted (and not as obvious as it might appear !) .

**REMARK** : If we use the alternate definition

$$\hat{S}^2 \quad \equiv \quad \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X})^2 \;,$$

then the Theorem becomes

$$\frac{n-1}{\sigma^2}\, \hat{S}^2 \;\text{ has the }\; \chi^2_{n-1} \;\text{ distribution .}$$

For normal random variables : $\dfrac{n-1}{\sigma^2}\,\hat{S}^2$ has the $\chi^2_{n-1}$ distribution

**EXAMPLE** : For a large shipment of light bulbs we are given that :

- The lifetime of the bulbs has a *normal distribution* .
- The mean lifetime $\mu$ is not given.
- The *standard deviation* is claimed to be $\sigma = \mathbf{100}$ hours.

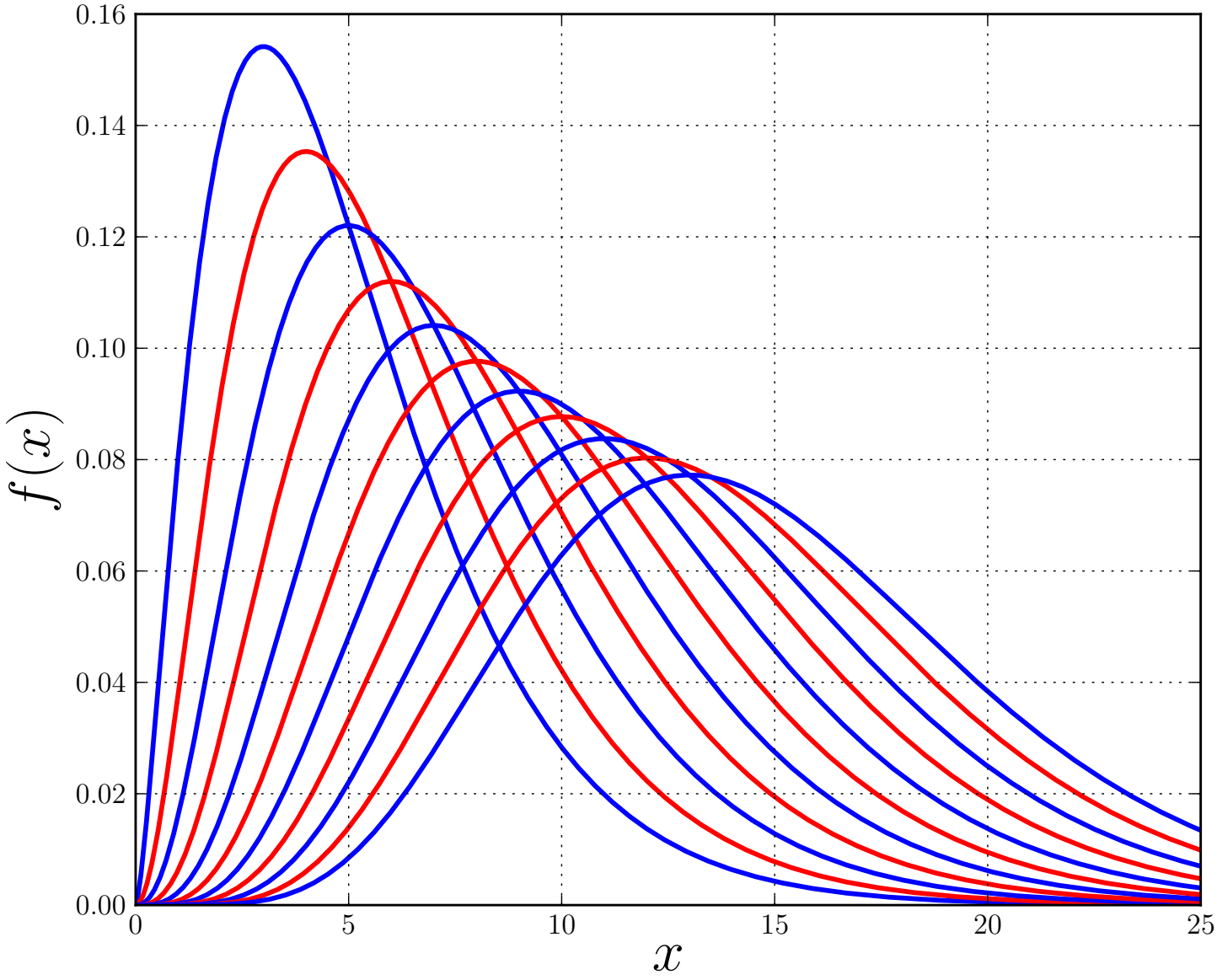
Suppose we test the lifetime of 16 bulbs. What is the probability that the sample standard deviation $\hat{S}$ satisfies $\hat{S} \geq \mathbf{129}$ hours ?

**SOLUTION** :

$$P(\hat{S} \geq 129) \;=\; P(\hat{S}^2 \geq 129^2) \;=\; P\Big(\frac{n-1}{\sigma^2}\,\hat{S}^2 \;\geq\; \frac{15}{100^2}\,129^2\Big)$$

$$\cong\; P\Big(\chi^2_{15} \;\geq\; 24.96\Big) \;\cong\; 5\,\% \quad \text{(from the } \chi^2 \text{ Table)} \; .$$

**QUESTION** : If $\hat{S} = 129$ then would you believe that $\sigma = 100$ ?




The Chi-Square *density* functions for $n = 5, 6, \cdots, 15$ .

(For *large* $n$ they look like *normal* density functions .)

**EXERCISE** :

In the preceding example, also compute

$$P\left(\chi^2_{15} \geq 24.96\right)$$

using the standard normal approximation .

**EXERCISE** :

Consider the same shipment of light bulbs :

- The lifetime of the bulbs has a *normal distribution* .
- The mean lifetime is not given.
- The *standard deviation* is claimed to be $\sigma = 100$ hours.

Suppose we test the lifetime of *only* 6 *bulbs* .

For what value of $s$ is $P(\hat{S} \leq s) = 5\ \%$ ?

**EXAMPLE** : For the data below from a *normal population* :

- Estimate the population standard deviation.
- Determine a 95 percent confidence interval for $\sigma$.

| | | | |
|---|---|---|---|
| -0.047 | 0.126 | -0.037 | 0.148 |
| 0.198 | 0.073 | -0.025 | -0.070 |
| -0.197 | -0.026 | -0.062 | -0.004 |
| -0.164 | 0.265 | -0.274 | 0.188 |

**SOLUTION** : ( start $\cdots$ ) : We find

$$\bar{X} \;=\; \frac{1}{n}\sum_{1}^{n} X_i \;=\; 0.00575\ ,$$

and

$$\hat{S}^2 \;=\; \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 \;=\; 0.02278\ .$$

**SOLUTION** : We have $n = 16$ , $\bar{X} = 0.00575$ , $\hat{S}^2 = 0.02278$ .

- Estimate the population standard deviation :

  **ANSWER** : $\sigma \cong \hat{S} = \sqrt{0.02278} = \mathbf{0.15095}$ .

- Compute a 95 percent confidence interval for $\sigma$ :

  **ANSWER** : From the *Chi-Square Table* :

  $$P(\chi^2_{15} \leq 6.26) = 0.025 \quad , \quad P(\chi^2_{15} > 27.49) = 0.025 \ .$$

  $$\frac{(n-1)\ \hat{S}^2}{\sigma^2} = 6.26 \quad \Rightarrow \quad \sigma^2 = \frac{(n-1)\hat{S}^2}{6.26} = \frac{15 \cdot 0.02278}{6.26} = 0.05458$$

  $$\frac{(n-1)\ \hat{S}^2}{\sigma^2} = 27.49 \quad \Rightarrow \quad \sigma^2 = \frac{(n-1)\hat{S}^2}{27.49} = \frac{15 \cdot 0.02278}{27.49} = 0.01223$$

  Thus the 95 % *confidence interval* for $\sigma$ is

  $$[\ \sqrt{0.01223}\ ,\ \sqrt{0.05458}\ ] = [\ \mathbf{0.106}\ ,\ \mathbf{0.234}\ ] \ .$$

## Samples from Finite Populations

Samples from a *finite population* can be taken

(1) *with replacement*

(2) *without replacement*

- In Case 1 the sample

$$X_1 \ , \ X_2 \ , \ \cdots \ , \ X_n \ ,$$

may contain the *same outcome* more than once.

- In Case 2 the outcomes are *distinct* .

- Case 2 arises, *e.g.*, when the experiment *destroys* the sample.

**EXAMPLE** :

Suppose a bag contains *three* balls, numbered 1, 2, and 3.

A *sample* of *two* balls is drawn at random from the bag.

Recall that

$$\bar{X} \quad \equiv \quad \frac{1}{n}\,(X_1 \;+\; X_2 \;+\; \cdots \;+\; X_n)\;.$$

$$S^2 \quad \equiv \quad \frac{1}{n}\sum_{k=1}^{n}\,(X_k - \bar{X})^2\;.$$

For both, sampling *with* and *without replacement* , compute

$$E[\bar{X}] \quad \text{and} \quad E[S^2]\;.$$

- *With replacement* : The possible samples are

  $(1,1)\,,\ (1,2)\,,\ (1,3)\,,\ (2,1)\,,\ (2,2)\,,\ (2,3)\,,\ (3,1)\,,\ (3,2)\,,\ (3,3)\,,$

  each with equal probability $\frac{1}{9}$ .

  The *sample means* $\bar{X}$ are

  $$1\ ,\ \frac{3}{2}\ ,\ 2\ ,\ \frac{3}{2}\ ,\ 2\ ,\ \frac{5}{2}\ ,\ 2\ ,\ \frac{5}{2}\ ,\ 3\ ,$$

  with

  $$E[\bar{X}] \ = \ \frac{1}{9}\,(1+\frac{3}{2}+2+\frac{3}{2}+2+\frac{5}{2}+2+\frac{5}{2}+3) \ = \ 2\,.$$

  The *sample variances* $S^2$ are

  $$0\ ,\ \frac{1}{4}\ ,\ 1\ ,\ \frac{1}{4}\ ,\ 0\ ,\ \frac{1}{4}\ ,\ 1\ ,\ \frac{1}{4}\ ,\ 0\,. \qquad (\text{ Check ! })$$

  with

  $$E[S^2] \ = \ \frac{1}{9}\,(0+\frac{1}{4}+1+\frac{1}{4}+0+\frac{1}{4}+1+\frac{1}{4}+0) \ = \ \frac{1}{3}\,.$$

- *Without replacement* : The possible samples are

$$(1,2)\ ,\ (1,3)\ ,\ (2,1)\ ,\ (2,3)\ ,\ (3,1)\ ,\ (3,2)\ ,$$

each with equal probability $\frac{1}{6}$ .

The *sample means* $\bar{X}$ are

$$\frac{3}{2}\ ,\ 2\ ,\ \frac{3}{2}\ ,\ \frac{5}{2}\ ,\ 2\ ,\ \frac{5}{2}\ ,$$

with *expected value*

$$E[\bar{X}] \ = \ \frac{1}{6}\,(\frac{3}{2} \ + \ 2 \ + \ \frac{3}{2} \ + \ \frac{5}{2} \ + \ 2 \ + \ \frac{5}{2}) \ \ = \ \ 2\ .$$

The *sample variances* $S^2$ are

$$\frac{1}{4}\ ,\ 1\ ,\ \frac{1}{4}\ ,\ \frac{1}{4}\ ,\ 1\ ,\ \frac{1}{4}\ .$$ ( Check ! )

with *expected value*

$$E[S^2] \ = \ \frac{1}{6}\,(\frac{1}{4} \ + \ 1 \ + \ \frac{1}{4} \ + \ \frac{1}{4} \ + \ 1 \ + \ \frac{1}{4}) \ \ = \ \ \frac{1}{2}\ .$$

**EXAMPLE** : ( continued $\cdots$ )

A bag contains *three* balls, numbered 1, 2, and 3.

A *sample* of *two* balls is drawn at random from the bag.

We have computed $E[\bar{X}]$ and $E[S^2]$ :

- *With* replacement : $E[\bar{X}] = 2$ , $E[S^2] = \frac{1}{3}$ ,
- *Without* replacement : $E[\bar{X}] = 2$ , $E[S^2] = \frac{1}{2}$ .

We also know the *population mean* and *variance* :

$$\mu = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} = 2 ,$$

$$\sigma^2 = (1-2)^2 \cdot \frac{1}{3} + (2-2)^2 \cdot \frac{1}{3} + (3-2)^2 \cdot \frac{1}{3} = \frac{2}{3} .$$

**EXAMPLE** : ( continued $\cdots$ )

We have computed :

- *Population* statistics : $\mu = 2 \quad , \quad \sigma^2 = \frac{2}{3}$ ,
- Sampling *with* replacement : $E[\bar{X}] = 2 \quad , \quad E[S^2] = \mathbf{\frac{1}{3}}$ ,
- Sampling *without* replacement : $E[\bar{X}] = 2 \quad , \quad E[S^2] = \frac{1}{2}$ .

According to the earlier Theorem

$$E[S^2] \quad = \quad (1 - \frac{1}{n})\, \sigma^2 \ .$$

In this example the *sample size* is $n = 2$ , thus

$$E[S^2] \quad = \quad (1 - \frac{1}{2})\, \sigma^2 \quad = \quad \mathbf{\frac{1}{3}} \ .$$

**NOTE** : $E[S^2]$ is *wrong* for sampling *without replacement* !

**QUESTION** :

Why is $E[S^2]$ *wrong* for sampling *without replacement* ?

**ANSWER** : Without replacement the outcomes $X_k$ of a sample

$$X_1 \ , \ X_2 \ , \ \cdots \ , \ X_n,$$

are *not independent* !

*In our example* , where $n = 2$ , and where the possible samples are

$$(1,2) \ , \ (1,3) \ , \ (2,1) \ , \ (2,3) \ , \ (3,1) \ , \ (3,2) \ ,$$

we have, *e.g.*,

$$P(X_2 = 1 \mid X_1 = 1) \ = \ 0 \quad , \quad P(X_2 = 1 \mid X_1 = 2) \ = \ \frac{1}{2} \ .$$

Thus $X_1$ and $X_2$ are *not independent* . ( Why not ? )

**NOTE** :

Let $N$ be the *population size* and $n$ the *sample size* .

Suppose $N$ is *very large* compared to $n$ .

For example, $n = 2$ , and the population is

$$\{ 1 , 2 , 3 , \cdots , N \} .$$

Then we still have

$$P(X_2 = 1 \mid X_1 = 1) \quad = \quad 0 ,$$

but for $k \neq 1$ we have

$$P(X_2 = k \mid X_1 = 1) \quad = \quad \frac{1}{N-1} .$$

One could say that $X_1$ and $X_2$ are "*almost independent*" . ( Why ? )

## The Sample Correlation Coefficient

Recall the *covariance* of random variables $X$ and $Y$ :

$$\sigma_{X,Y} \equiv Cov(X,Y) = E[\,(X-\mu_X)\,(Y-\mu_Y)\,] = E[XY] - E[X]\,E[Y]\,.$$

It is often better to use a *scaled* version, the *correlation coefficient*

$$\rho_{X,Y} \equiv \frac{\sigma_{X,Y}}{\sigma_X\ \sigma_Y}\ ,$$

where $\sigma_X$ and $\sigma_Y$ are the standard deviation of $X$ and $Y$ .

We have

- $|\,\sigma_{X,Y}\,| \leq \sigma_X\ \sigma_Y\ ,$ (the *Cauchy-Schwartz inequality* )
- Thus $|\,\rho_{X,Y}\,| \leq 1\ ,$ ( Why ? )
- If $X$ and $Y$ are independent then $\rho_{X,Y} = 0\ .$ ( Why ? )

Similarly, the *sample correlation coefficient* of a data set

$$\{ (X_i, Y_i) \}_{i=1}^{N} ,$$

is defined as

$$R_{X,Y} \equiv \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2}\,\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}} ;$$

for which we have another version of the *Cauchy-Schwartz inequality* :

$$| R_{X,Y} | \leq 1 .$$

Like the covariance, $R_{X,Y}$ measures "*concordance*" of $X$ and $Y$ :

- If $X_i > \bar{X}$ when $Y_i > \bar{Y}$ and $X_i < \bar{X}$ when $Y_i < \bar{Y}$ then
$$R_{X,Y} > 0 .$$
- If $X_i > \bar{X}$ when $Y_i < \bar{Y}$ and $X_i < \bar{X}$ when $Y_i > \bar{Y}$ then
$$R_{X,Y} < 0 .$$

The *sample correlation coefficient*

$$R_{X,Y} \equiv \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}} .$$

can also be used to *test for linearity* of the data.

In fact,

- If $|R_{X,Y}| = 1$ then $X$ and $Y$ are related *linearly* .

Specifically,

- If $R_{X,Y} = 1$ then $Y_i = cX_i + d$, for constants $c, d$, with $c > 0$ .
- If $R_{X,Y} = -1$ then $Y_i = cX_i + d$, for constants $c, d$, with $c < 0$ .

Also,

- If $|R_{X,Y}| \cong 1$ then $X$ and $Y$ are *almost linear* .

**EXAMPLE** :

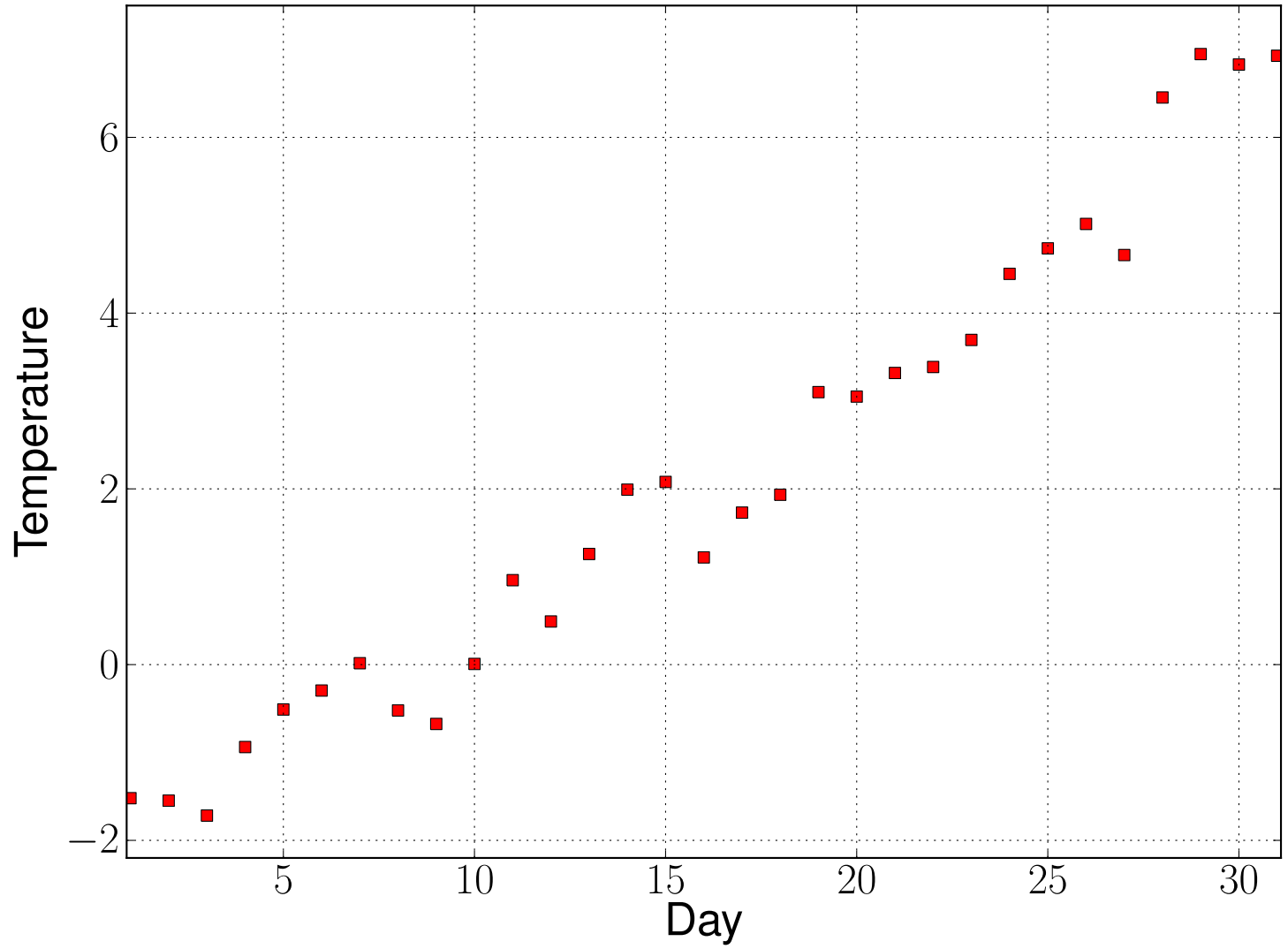
- Consider the *average daily high temperature* in Montreal in March.
- The Table shows these averages, taken over a number of years :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | -1.52 | 8 | -0.52 | 15 | 2.08 | 22 | 3.39 | 29 | 6.95 |
| 2 | -1.55 | 9 | -0.67 | 16 | 1.22 | 23 | 3.69 | 30 | 6.83 |
| 3 | -1.72 | 10 | 0.01 | 17 | 1.73 | 24 | 4.45 | 31 | 6.93 |
| 4 | -0.94 | 11 | 0.96 | 18 | 1.93 | 25 | 4.74 | | |
| 5 | -0.51 | 12 | 0.49 | 19 | 3.10 | 26 | 5.01 | | |
| 6 | -0.29 | 13 | 1.26 | 20 | 3.05 | 27 | 4.66 | | |
| 7 | 0.02 | 14 | 1.99 | 21 | 3.32 | 28 | 6.45 | | |

Average daily high temperature in Montreal in March : 1943-2014 .

( Source : `http://climate.weather.gc.ca/` )

These data have *sample correlation coefficient* $R_{X,Y} = \mathbf{0.98}$ .




A *scatter diagram* showing the average daily high temperature.

The sample correlation coefficient is $R_{X,Y} = \mathbf{0.98}$

**EXERCISE** :

- The Table below shows class attendance and course grade/100.
- The attendance was sampled in 18 sessions.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | 47 | 13 | 43 | 15 | 70 | 17 | 72 | 18 | 96 | 14 | 61 | 5 | 25 | 17 | 74 |
| 16 | 85 | 13 | 82 | 16 | 67 | 17 | 91 | 16 | 71 | 16 | 50 | 14 | 77 | 12 | 68 |
| 8 | 62 | 13 | 71 | 12 | 56 | 15 | 81 | 16 | 69 | 18 | 93 | 18 | 77 | 17 | 48 |
| 14 | 82 | 17 | 66 | 16 | 91 | 17 | 67 | 7 | 43 | 15 | 86 | 18 | 85 | 17 | 84 |
| 11 | 43 | 17 | 66 | 18 | 57 | 18 | 74 | 13 | 73 | 15 | 74 | 18 | 73 | 17 | 71 |
| 14 | 69 | 15 | 85 | 17 | 79 | 18 | 84 | 17 | 70 | 15 | 55 | 14 | 75 | 15 | 61 |
| 16 | 61 | 4 | 46 | 18 | 70 | 0 | 29 | 17 | 82 | 18 | 82 | 16 | 82 | 14 | 68 |
| 9 | 84 | 15 | 91 | 15 | 77 | 16 | 75 | | | | | | | | |

Class attendance - Course grade

- Draw a *scatter diagram* showing the data.
- Determine the *sample correlation coefficient* .
- Any *conclusions* ?

## Maximum Likelihood Estimators

**EXAMPLE** :

Suppose a random variable has a *normal distribution* with mean 0 .

Thus the density function is

$$f(x) \quad = \quad \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{1}{2}x^2/\sigma^2} \ .$$

- Suppose we don't know $\sigma$ (the *population standard deviation*).
- How can we *estimate* $\sigma$ from observed data ?
- (We want a *formula* for estimating $\sigma$ .)
- Don't we already have such a formula ?

**EXAMPLE** : ( continued $\cdots$ )

We know we can *estimate* $\sigma^2$ by the *sample variance*

$$S^2 \equiv \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2 \ .$$

In fact, we have proved that

$$E[S^2] = (1 - \frac{1}{n})\,\sigma^2 \ .$$

- Thus, we can call $S^2$ an *estimator* of $\sigma^2$ .
- The "maximum likelihood procedure" *derives* such estimators.

The *maximum likelihood procedure* is the following :

Let

$$X_1 \ , \ X_2 \ , \ \cdots \ , \ X_n \ ,$$

be

*independent, identically distributed* ,

each having

density function $f(x \ ; \ \sigma)$ ,

with *unknown parameter* $\sigma$ .

By *independence* , the *joint density function* is

$$f(x_1, x_2, \ \cdots \ , x_n \ ; \ \sigma) \ = \ f(x_1; \sigma) \ f(x_2; \sigma) \ \cdots \ f(x_n; \sigma) \ ,$$

**DEFINITION** : The *maximum likelihood estimate* $\hat{\sigma}$ is

the value of $\sigma$ that *maximizes* $f(x_1, x_2, \ \cdots \ , x_n \ ; \ \sigma)$ .

**NOTE** : $\hat{\sigma}$ will be a *function* of $x_1, x_2, \ \cdots \ , x_n$ .

**EXAMPLE** : For our *normal distribution* with mean 0 we have

$$f(x_1, x_2, \cdots, x_n \,;\, \sigma) \;=\; \frac{e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{n} x_k^2}}{(\sqrt{2\pi}\;\sigma)^n} \,. \qquad (\text{ Why ? })$$

To find the maximum (with respect to $\sigma$ ) we set

$$\frac{d}{d\sigma}\, f(x_1, x_2, \cdots, x_n \,;\, \sigma) \;=\; 0 \,, \qquad (\text{ by Calculus ! })$$

or, *equivalently* , we set

$$\frac{d}{d\sigma}\, \log\!\left(\frac{e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{n} x_k^2}}{\sigma^n}\right) \;=\; 0 \,. \qquad (\text{ Why } \textit{equivalent} \text{ ? })$$

Taking the (natural) logarithm gives

$$\frac{d}{d\sigma}\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{n} x_k^2 \;-\; n\,\log\sigma\right) \;=\; 0 \,.$$

**EXAMPLE** : ( continued $\cdots$ )

We had

$$\frac{d}{d\sigma}\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{n} x_k^2 \;-\; n\,\log\,\sigma\right) \;=\; 0\,.$$

Taking the derivative gives

$$\frac{\sum_{k=1}^{n} x_k^2}{\sigma^3} \;-\; \frac{n}{\sigma} \;=\; 0\,,$$

from which

$$\hat{\sigma}^2 \;=\; \frac{1}{n}\sum_{k=1}^{n} x_k^2\,.$$

Thus we have derived the *maximum likelihood estimate*

$$\hat{\sigma} \;=\; \frac{1}{\sqrt{n}}\left(\sum_{k=1}^{n} X_k^2\right)^{\frac{1}{2}}\,.$$

**EXERCISE** :

Suppose a random variable has the *general normal density function*

$$f(x\ ;\ \mu, \sigma) \quad = \quad \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}\ ,$$

with *unknown* mean $\mu$ *and unknown* standard deviation $\sigma$ .

Derive *maximum likelihood estimators* for *both* $\mu$ and $\sigma$ as follows :

For the *joint density function*

$$f(x_1, x_2, \cdots, x_n; \mu, \sigma) \ = \ f(x_1; \mu, \sigma)\ f(x_2; \mu, \sigma) \cdots f(x_n; \mu, \sigma)\ ,$$

- Take the log of $f(x_1, x_2, \ \cdots\ , x_n\ ;\ \mu, \sigma)$ .
- Set the *partial derivative* w.r.t. $\mu$ equal to zero.
- Set the *partial derivative* w.r.t. $\sigma$ equal to zero.
- Solve these two equations for $\hat{\mu}$ and $\hat{\sigma}$ .

**EXERCISE** : ( continued $\cdots$ )

The *maximum likelihood estimators* should turn out to be

$$\hat{\mu} \;=\; \frac{1}{n}\sum_{k=1}^{n} X_k \;,$$

$$\hat{\sigma} \;=\; \frac{1}{\sqrt{n}}\left(\sum_{k=1}^{n}(X_k-\bar{X})^2\right)^{\frac{1}{2}} \;,$$

that is,

$$\hat{\mu} \;=\; \bar{X} \;, \qquad (\text{the } \textit{sample mean}) \;,$$

$$\hat{\sigma} \;=\; S \quad (\text{the } \textit{sample standard deviation}) \;.$$

**NOTE** :

- Earlier we simply *defined* the *sample variance* as

$$S^2 \quad = \quad \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2 \ .$$

- Then we proved that, in general,

$$E[S^2] \quad = \quad (1 - \frac{1}{n})\ \sigma^2 \quad \cong \quad \sigma^2.$$

- In the preceding exercise we *derived* the estimator for $\sigma^2$ !

- (But we did so *specifically* for the general normal distribution.)

**EXERCISE** :

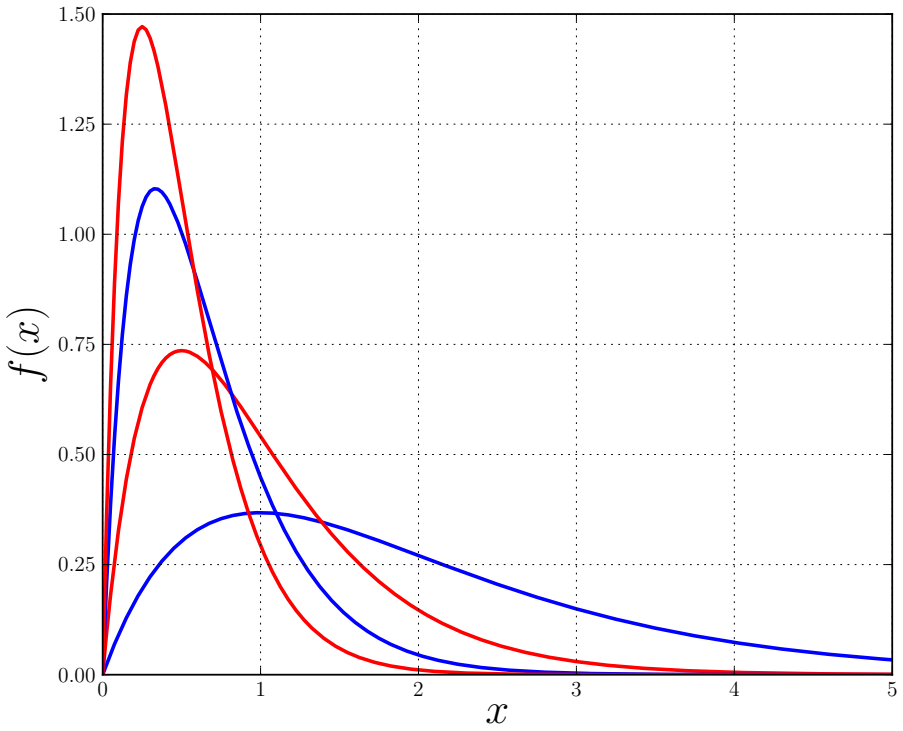
A random variable has the standard *exponential distribution* with density function

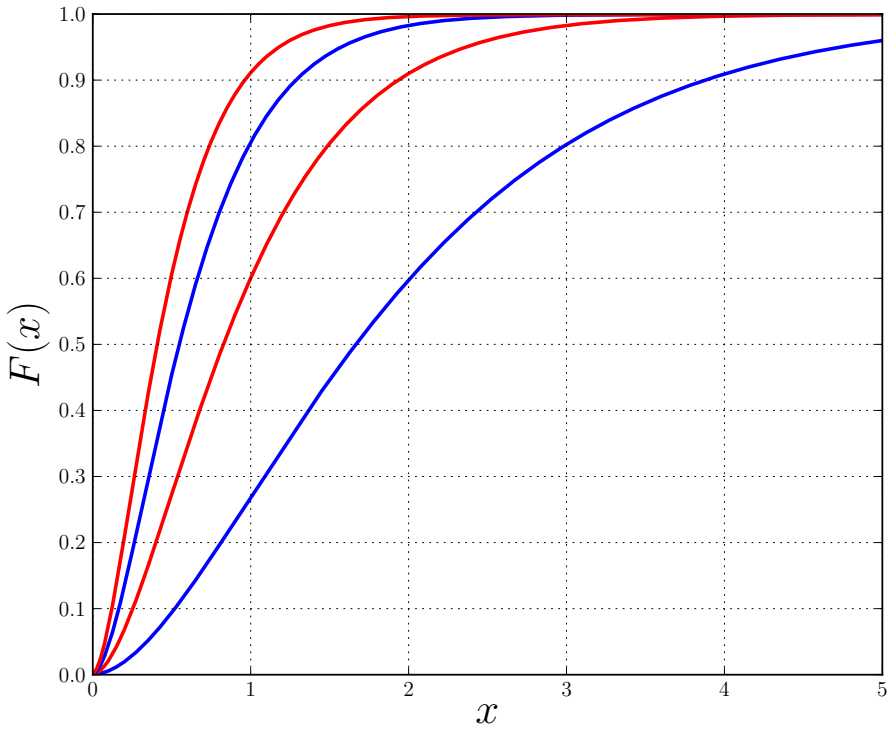
$$f(x\ ;\ \lambda)\ =\ \begin{cases} \lambda e^{-\lambda x}\ , & x > 0 \\ \\ 0\ , & x \leq 0 \end{cases}$$

- Suppose we don't know $\lambda$ .
- Derive the *maximum likelihood estimator* of $\lambda$ .
- (Can you guess what the formula will be ?)

**EXAMPLE** : Consider the *special* exponential density function

$$f(x\ ;\ \lambda)\ =\ \begin{cases} \lambda^2 x e^{-\lambda x}\ , & x > 0 \\ \\ 0\ , & x \leq 0 \end{cases}$$




Density and distribution functions for $\lambda = 1\ ,\ 2\ ,\ 3\ ,\ 4\ .$

**EXAMPLE** : ( continued $\cdots$ )

For the *maximum likelihood estimator* of $\lambda$ , we have

$$f(x\ ;\ \lambda) \quad = \quad \lambda^2\ x\ e^{-\lambda x}\ , \qquad \text{for } x > 0\ ,$$

so, assuming independence, the *joint density function* is

$$f(x_1, x_2,\ \cdots\ , x_n\ ;\ \lambda) \quad = \quad \lambda^{2n}\ x_1 x_2\ \cdots\ x_n\ e^{-\lambda(x_1+x_2+\ \cdots\ +x_n)}\ .$$

To find the *maximum* (with respect to $\lambda$ ) we set

$$\frac{d}{d\lambda}\ \log\Big(\lambda^{2n}\ \ x_1 x_2\ \cdots\ x_n\ \ e^{-\lambda(x_1+x_2+\ \cdots\ +x_n)}\ \Big) \quad = \quad 0\ .$$

Taking the logarithm gives

$$\frac{d}{d\lambda}\ \Big(2n\ \log\ \lambda\ \ +\ \ \sum_{k=1}^{n}\ \log\ x_k\ \ -\ \ \lambda \sum_{k=1}^{n}\ x_k\Big) \ = \ 0\ .$$

**EXAMPLE** : ( continued $\cdots$ )

We had

$$\frac{d}{d\lambda}\left(2n\,\log\,\lambda \;+\; \sum_{k=1}^{n}\,\log\,x_k \;-\; \lambda\sum_{k=1}^{n}\,x_k\right) \;=\; 0\;.$$

Differentiating gives

$$\frac{2n}{\lambda} \;-\; \sum_{k=1}^{n}\,x_k \;=\; 0\;,$$

from which

$$\hat{\lambda} \;=\; \frac{2n}{\sum_{k=1}^{n}\,x_k}\;.$$

Thus we have derived the *maximum likelihood estimate*

$$\hat{\lambda} \;=\; \frac{2n}{\sum_{k=1}^{n}\,X_k} \;=\; \frac{2}{\bar{X}}\;.$$

**NOTE** : This result suggests that perhaps $E[X] \;=\; 2/\lambda\,.$ ( Why ? )

**EXERCISE** :

For the special exponential density function in the preceding example,

$$f(x\ ;\ \lambda) \ = \ \begin{cases} \lambda^2 x e^{-\lambda x}\ , & x > 0 \\ \\ 0\ , & x \leq 0 \end{cases}$$

- Verify that

$$\int_0^\infty f(x\ ;\ \lambda)\ dx \quad = \quad 1\ .$$

- Also compute

$$E[X] \quad = \quad \int_0^\infty x\ f(x\ ;\ \lambda)\ dx$$

- Is it indeed true that

$$E[X] \quad = \quad \frac{2}{\lambda} \qquad ?$$

**NOTE** :

- Maximum likelihood estimates also work in the *discrete case* .
- In such case we maximize the *probability mass function* .

**EXAMPLE** :

Find the maximum likelihood estimator of $p$ in the *Bernoulli trial*

$$\begin{aligned} P(X=1) &= p\ , \\ P(X=0) &= 1-p\ . \end{aligned}$$

**SOLUTION** : We can write

$$P(x\ ;\ p) \equiv P(X=x) = p^x\ (1-p)^{1-x}\ , \quad (x=0,1) \qquad (!)$$

so, assuming *independence* , the *joint probability mass function* is

$$\begin{aligned} P(x_1, x_2, \cdots, x_n; p) &= p^{x_1}(1-p)^{1-x_1}\ p^{x_2}(1-p)^{1-x_2}\ \cdots\ p^{x_n}(1-p)^{1-x_n} \\ &= p^{\sum_{k=1}^n x_k}\ \cdot\ (1-p)^n\ \cdot\ (1-p)^{-\sum_{k=1}^n x_k}\ . \end{aligned}$$

**EXAMPLE** : ( continued $\cdots$ )

We found

$$P(x_1, x_1, \ \cdots \ , x_n; p) \quad = \quad p^{\sum_{k=1}^{n} x_k} \ \cdot \ (1-p)^n \ \cdot \ (1-p)^{-\sum_{k=1}^{n} x_k} \quad .$$

To find the *maximum* (with respect to $p$ ) we set

$$\frac{d}{dp} \log\Big( p^{\sum_{k=1}^{n} x_k} \ \cdot \ (1-p)^n \ \cdot \ (1-p)^{-\sum_{k=1}^{n} x_k} \Big) \quad = \quad 0 \ .$$

Taking the logarithm gives

$$\frac{d}{dp} \left( \log p \sum_{k=1}^{n} x_k \quad + \quad n \log(1-p) \quad - \quad \log (1-p) \sum_{k=1}^{n} x_k \right) \ = \ 0 \ .$$

Differentiating gives

$$\frac{1}{p} \sum_{k=1}^{n} x_k \quad - \quad \frac{n}{1-p} \quad + \quad \frac{1}{1-p} \sum_{k=1}^{n} x_k \quad = \quad 0 \ .$$

**EXAMPLE** : ( continued $\cdots$ )

We found

$$\frac{1}{p}\sum_{k=1}^{n} x_k \quad - \quad \frac{n}{1-p} \quad + \quad \frac{1}{1-p}\sum_{k=1}^{n} x_k \quad = \quad 0 \ ,$$

from which

$$\left( \frac{1}{p} \ + \ \frac{1}{1-p} \right) \sum_{k=1}^{n} x_k \quad = \quad \frac{n}{1-p} \ .$$

Multiplying by $1-p$ gives

$$\left( \frac{1-p}{p} + 1 \right) \sum_{k=1}^{n} x_k \quad = \quad \frac{1}{p}\sum_{k=1}^{n} x_k \quad = \quad n \ ,$$

from which we obtain the *maximum likelihood estimator*

$$\hat{p} \quad = \quad \frac{\sum_{k=1}^{n} X_k}{n} \quad \equiv \quad \bar{X} \ . \qquad \text{( Surprise ? )}$$

**EXERCISE** :

Consider the *Binomial* probability mass function

$$P(x\ ;\ p) \quad \equiv \quad P(X = x) \quad = \quad \binom{N}{x} \cdot p^x \cdot (1-p)^{N-x}\ ,$$

where $x$ is an integer, $(0 \leq x \leq N)$ .

- What is the *joint probability mass function* $P(x_1, x_2, \ \cdots\ , x_n; p)$ ?
- ( Be sure to distinguish between $N$ and $n$ ! )
- Determine the *maximum likelihood estimator* $\hat{p}$ of $p$ .
- ( Can you guess what $\hat{p}$ will be ? )

## Hypothesis Testing

- Often we want to decide whether a *hypothesis* is True or False.
- To do so we gather data, *i.e.*, a *sample* .
- A typical hypothesis is that a random variable *has a given mean* .
- Based on the data we want to *accept* or *reject* the hypothesis.
- To illustrate concepts we consider an example in detail.

**EXAMPLE** :

We consider ordering a large shipment of 50 watt light bulbs.

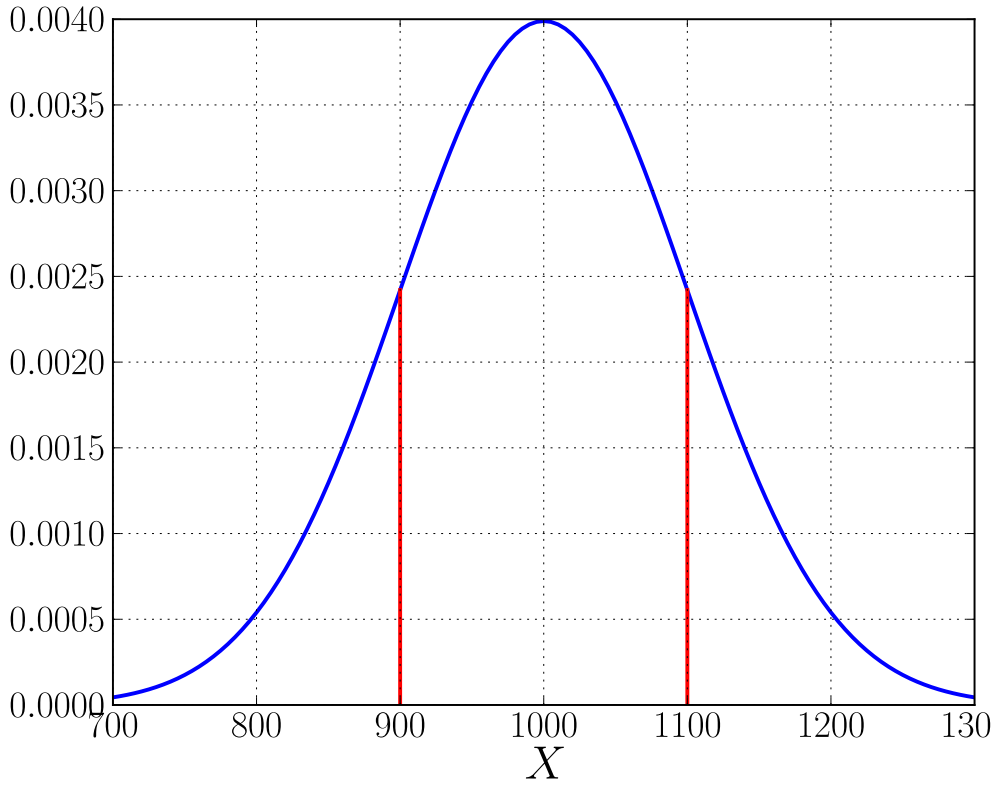
The manufacturer *claims* that :

- The lifetime of the bulbs has a *normal distribution* .
- The mean lifetime is $\mu = 1000$ hours.
- The standard deviation is $\sigma = 100$ hours.

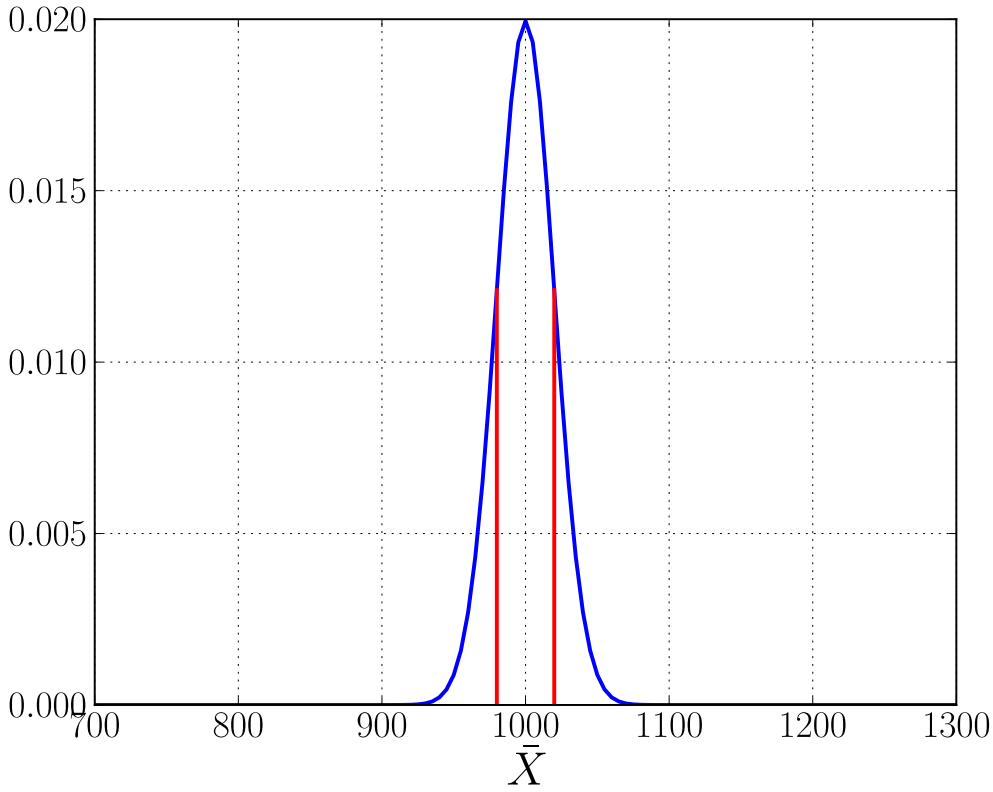
We want to *test the hypothesis that* $\mu = 1000$ .

We *assume* that :

- The lifetime of the bulbs has indeed a *normal distribution* .
- The standard deviation is indeed $\sigma = 100$ hours.
- We *test* the lifetime of a sample of 25 *bulbs* .




Density function of $X$ ,
also indicating $\mu \pm \sigma_X$ ,
$(\mu_X = 1000\ ,\ \ \sigma_X = 100)$ .




Density function of $\bar{X}$ $(n = 25)$ ,
also indicating $\mu_{\bar{X}} \pm \sigma_{\bar{X}}$ ,
$(\mu_{\bar{X}} = 1000\ ,\ \ \sigma_{\bar{X}} = 20\ )$.

**EXAMPLE** : ( continued $\cdots$ )

We *test* a *sample* of 25 light bulbs :

- We find the *sample* average lifetime is $\bar{X} = 960$ hours.

- Do we *accept* the hypothesis that $\mu = 1000$ hours ?

Using the standard normal Table we have the *one-sided probability*

$$P(\bar{X} \leq 960) \quad = \quad \Phi\Big(\frac{960 - 1000}{100/\sqrt{25}}\Big) \quad = \quad \Phi(-2.0) \quad = \quad 2.28\ \% \ ,$$

(assuming that the average lifetime is indeed 1000 hours).

- Would you *accept* the hypothesis that $\mu = 1000$ ?

- Would you *accept* (and pay for !) the shipment ?

**EXAMPLE** : ( continued $\cdots$ )

We *test* a *sample* of 25 light bulbs :

- Suppose instead the *sample* average lifetime is $\bar{X} = 1040$ hours.

- Do we *accept* that $\mu = 1000$ hours ?

Using the standard normal Table we have *one-sided probability*

$$P(\bar{X} \geq 1040) \;=\; 1 - \Phi\Big(\frac{1040 - 1000}{100/\sqrt{25}}\Big) \;=\; 1 - \Phi(2) \;=\; \Phi(-2) \;=\; 2.28\,\%\,,$$

(assuming again that the average lifetime is indeed 1000 hours).

- Would you *accept the hypothesis* that that the mean is 1000 hours ?

- Would you *accept* the shipment ? ( ! )

**EXAMPLE** : ( continued $\cdots$ )

Suppose that we *accept the hypothesis* that $\mu = 1000$ if

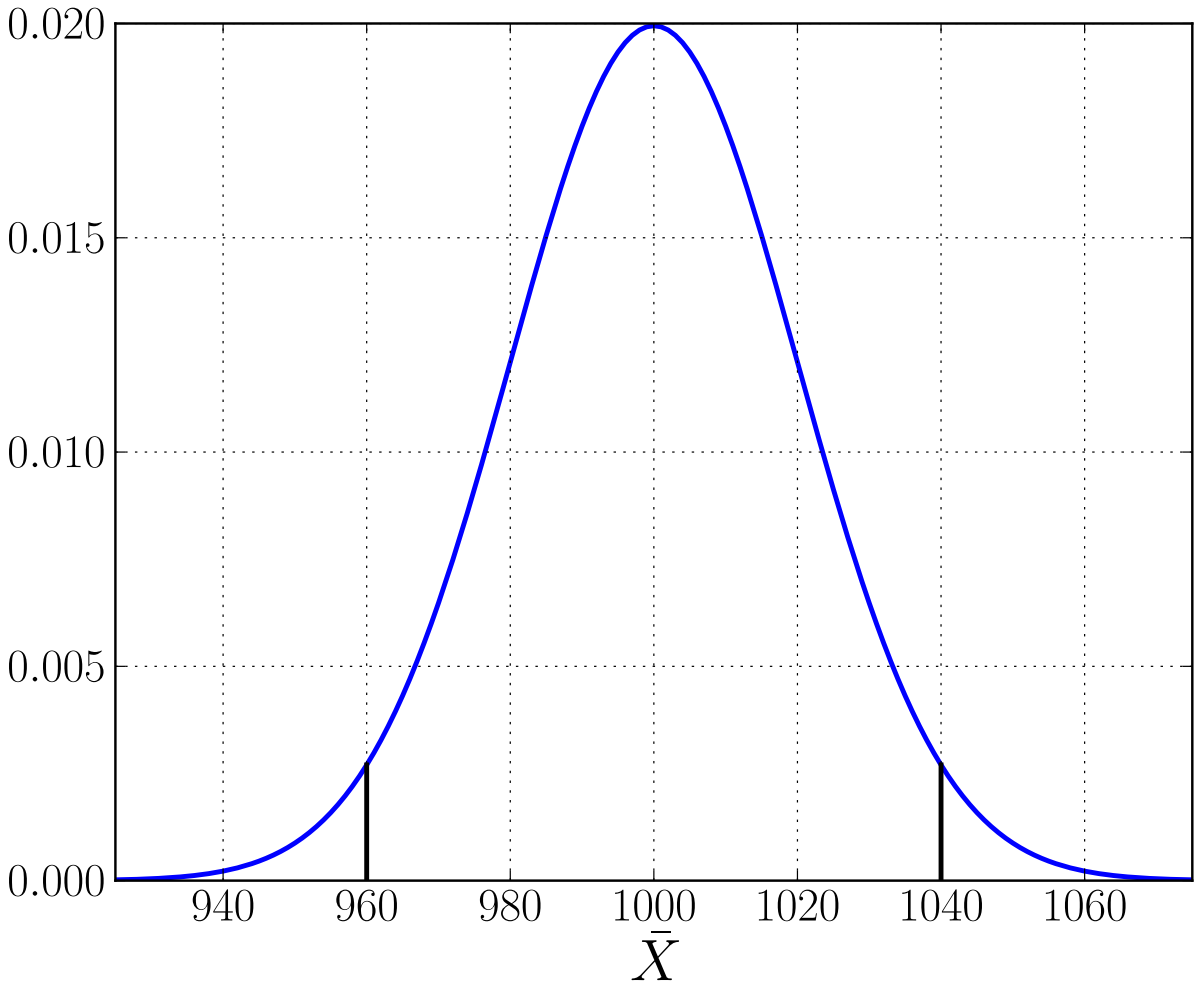
$$960 \;\leq\; \bar{X} \;\leq\; 1040 \,.$$

Thus, if indeed $\mu = 1000$ , we *accept* the hypothesis with probability

$$P(\;\; |\, \bar{X} - 1000 \,| \leq 40 \,) \;\; = \;\; 1 - 2\Phi\Big(\frac{960 - 1000}{100/\sqrt{25}}\Big) \;\; = \;\; 1 - 2\Phi(-2) \;\; \cong \;\; 95\,\% \,,$$

and we *reject* the hypothesis with probability

$$P(\;\; |\, \bar{X} - 1000 \,| \;\; \geq \;\; 40 \,) \;\; = \;\; 100\,\% \;-\; 95\,\% \;\; = \;\; 5\,\% \,.$$




Density function of $\bar{X}$ $(n = 25)$ , with $\mu = \mu_{\bar{X}} = 1000$ , $\sigma_{\bar{X}} = 20$ ,

$$P(960 \leq \bar{X} \leq 1040) \cong 95\%$$

**EXAMPLE** : ( continued $\cdots$ )

What is the probability of

*acceptance* of the hypothesis if $\mu$ is different from 1000 ?

- If the *actual mean* is $\mu = 980$ , then acceptance has probability

$$P(960 \leq \bar{X} \leq 1040) = \Phi(\frac{1040-980}{100/\sqrt{25}}) - \Phi(\frac{960-980}{100/\sqrt{25}})$$

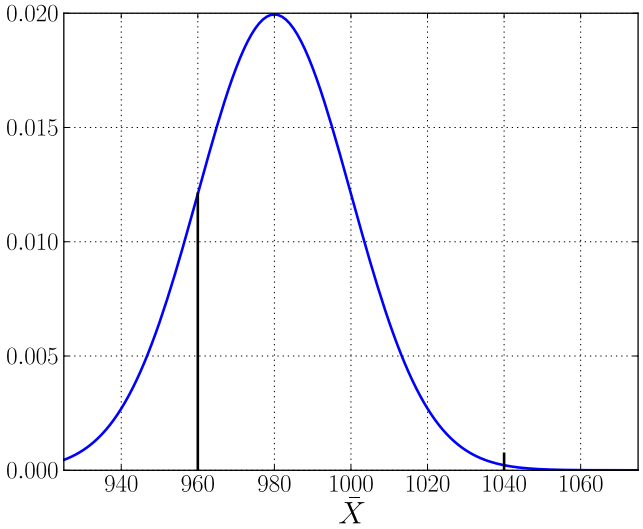
$$= \Phi(3) - \Phi(-1) = 1 - \Phi(-3) - \Phi(-1)$$

$$(1 - 0.0013) - 0.1587 = 84\ \% \ .$$

- If the *actual mean* is $\mu = 1040$ , then acceptance has probability

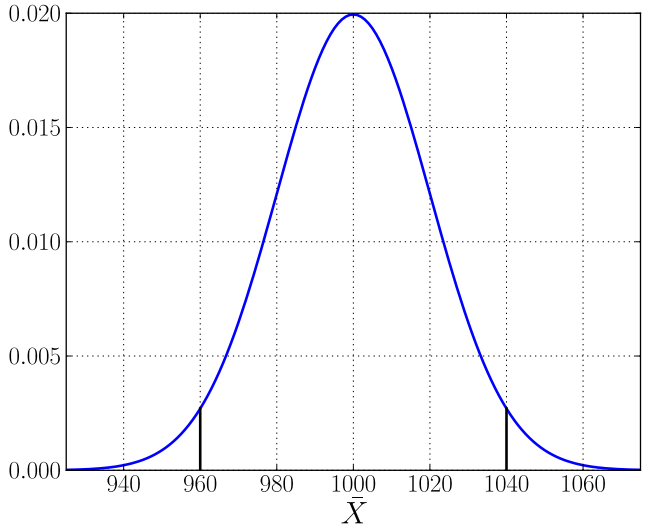
$$P(960 \leq \bar{X} \leq 1040) = \Phi(\frac{1040-1040}{100/\sqrt{25}}) - \Phi(\frac{960-1040}{100/\sqrt{25}})$$

$$= \Phi(0) - \Phi(-4) \cong 50\ \% \ .$$



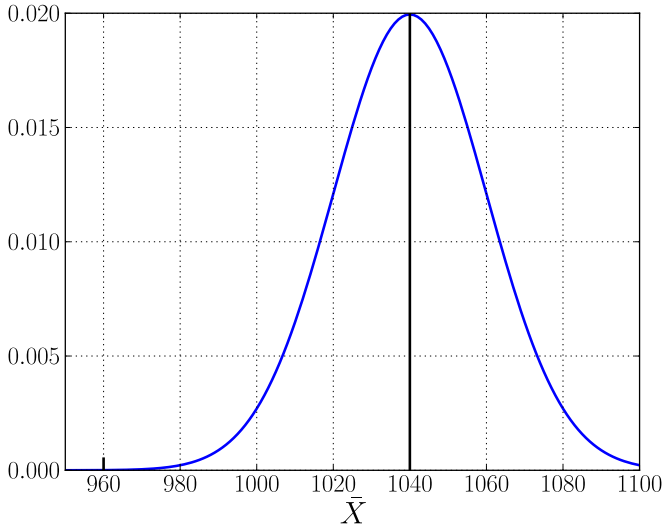

$\mu = \mu_{\bar{X}} = 980$

$P(accept) = 84\%$




$\mu = \mu_{\bar{X}} = 1000$

$P(accept) = 95\%$




$\mu = \mu_{\bar{X}} = 1040$

$P(accept) = 50\%$

Density functions of $\bar{X}$ : $n = 25$ , $\sigma_{\bar{X}} = 20$

**QUESTION 1** : How does $P(accept)$ change when we "slide" the density function of $\bar{X}$ along the $\bar{X}$-axis , *i.e.*, when $\mu$ changes ?

**QUESTION 2** : What is the effect of increasing the *sample size* $n$ ?

**EXAMPLE** :

Now suppose there are *two lots* of light bulbs :

- Lot 1 : Light bulbs with mean life time $\mu_1 = 1000$ hours,
- Lot 2 : Light bulbs with mean life time $\mu_2 = 1100$ hours.

We want to *decide* which lot our sample of 25 bulbs is from.

Consider the *decision criterion* $\hat{x}$ , where $1000 \leq \hat{x} \leq 1100$ :

- If $\bar{X} \leq \hat{x}$ then the sample is from Lot 1 .
- If $\bar{X} > \hat{x}$ then the sample is from Lot 2 .

There are *two hypotheses* :

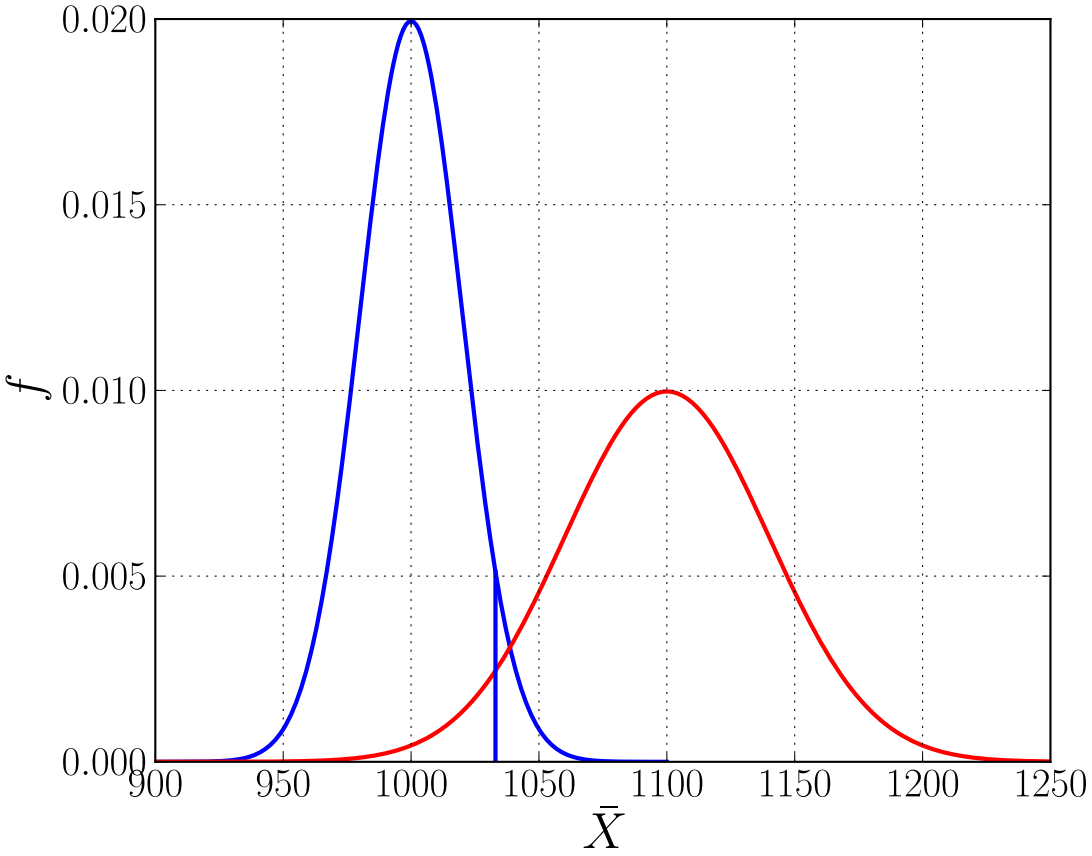
- H1 : The sample is from Lot 1 $(\mu_1 = 1000)$ .
- H2 : The sample is from Lot 2 $(\mu_2 = 1100)$ .

We can make *two types of errors* :

- Type 1 error : Accept H2 when H1 is True ,
- Type 2 error : Accept H1 when H2 is True ,

which happen when, for given *decision criterion* $\hat{x}$ ,

- Type 1 error : If $\bar{X} > \hat{x}$ and the sample is from Lot 1 .
- Type 2 error : If $\bar{X} \leq \hat{x}$ and the sample is from Lot 2 .




The density functions of $\bar{X}$ $(n = 25)$ , also indicating $\hat{x}$ .

*blue* : $(\mu_1, \sigma_1) = (1000, 100)$ , *red* : $(\mu_2, \sigma_2) = (1100, 200)$ .

Type 1 error : area under the *blue* curve, to the right of $\hat{x}$ .

Type 2 error : area under the *red* curve, to the left of $\hat{x}$ .

**QUESTION** : What is the effect of moving $\hat{x}$ on these errors ?

**RECALL** :

- Type 1 error : If $\bar{X} > \hat{x}$ and the sample is from Lot 1 .
- Type 2 error : If $\bar{X} \leq \hat{x}$ and the sample is from Lot 2 .

These errors occur with *probability*

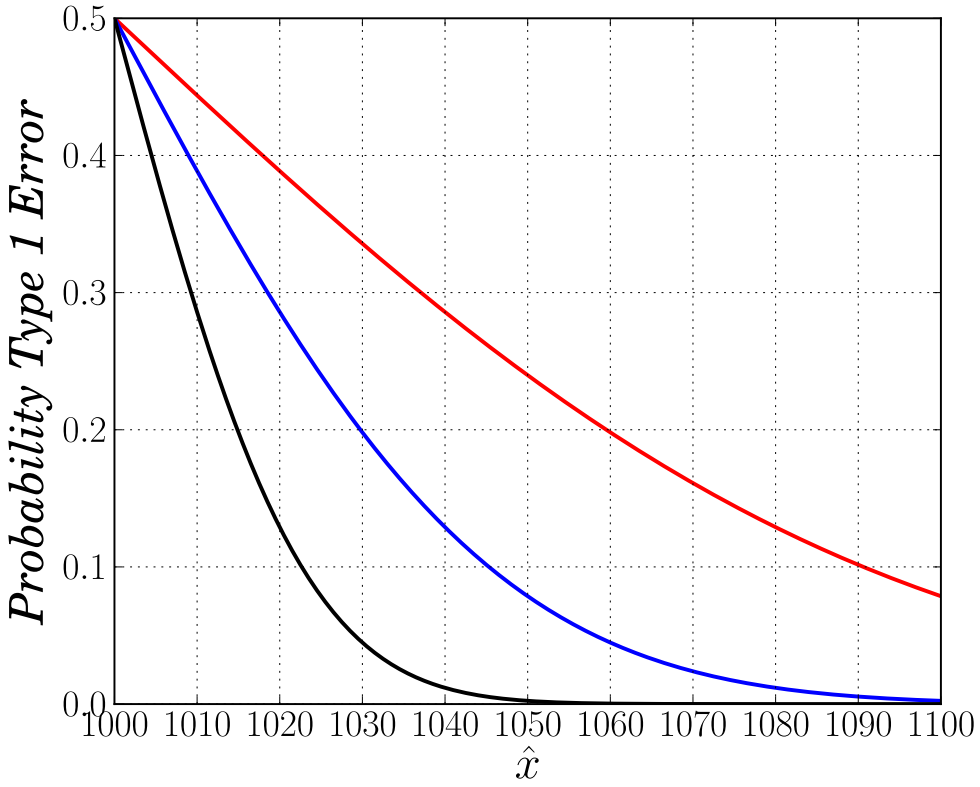
- Type 1 error : $P(\bar{X} \geq \hat{x} \mid \mu = \mu_1 \equiv 1000)$ .
- Type 2 error : $P(\bar{X} \leq \hat{x} \mid \mu = \mu_2 \equiv 1100)$ .

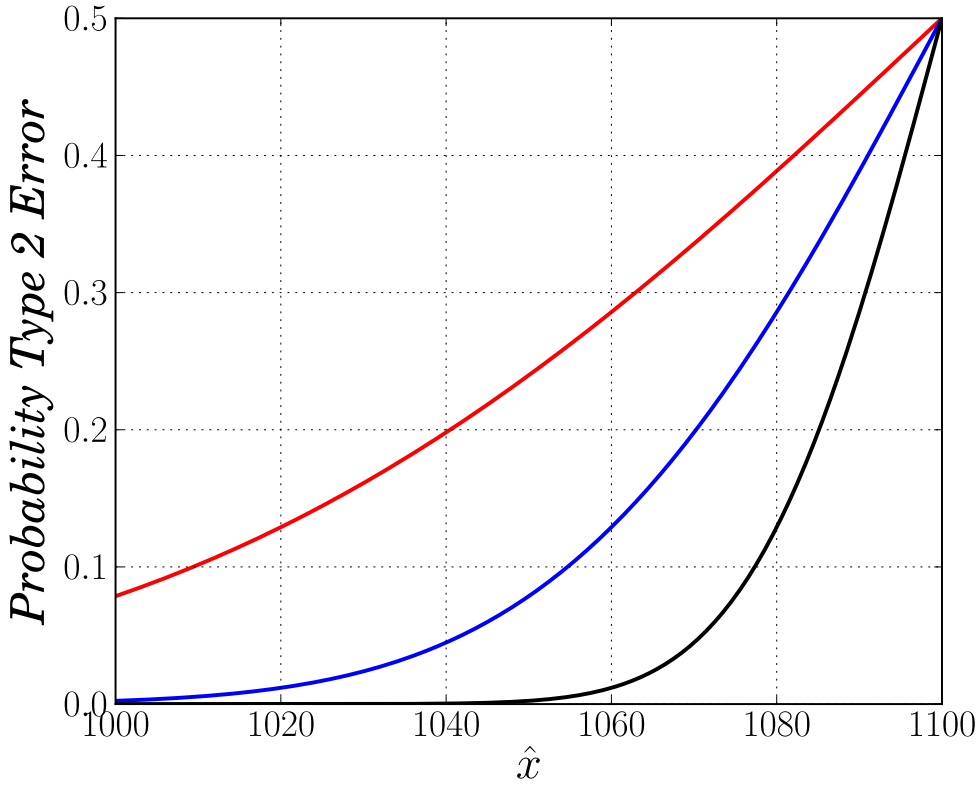
We should have, for the (rather bad) choice $\hat{x} = 1000$ ,

- Type 1 error : $P(\bar{X} \geq 1000 \mid \mu = \mu_1 \equiv 1000) = 0.5$ .

and for the (equally bad) choice $\hat{x} = 1100$ ,

- Type 2 error : $P(\bar{X} \leq 1100 \mid \mu = \mu_2 \equiv 1100) = 0.5$ .



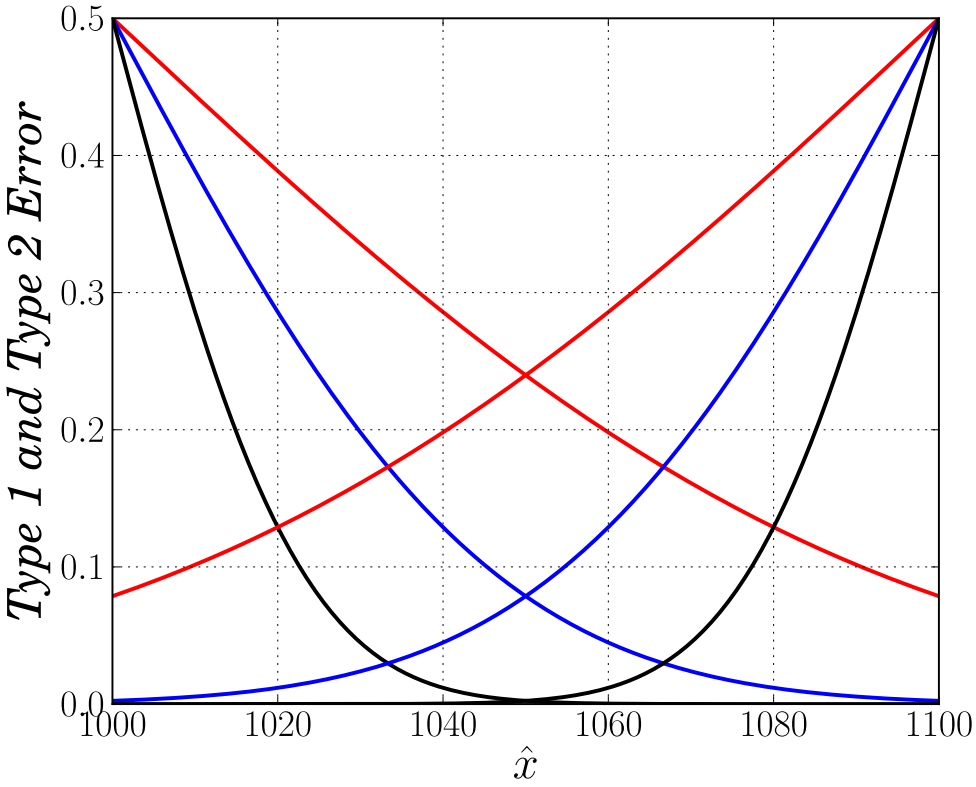

Probability of Type 1 error vs. $\hat{x}$

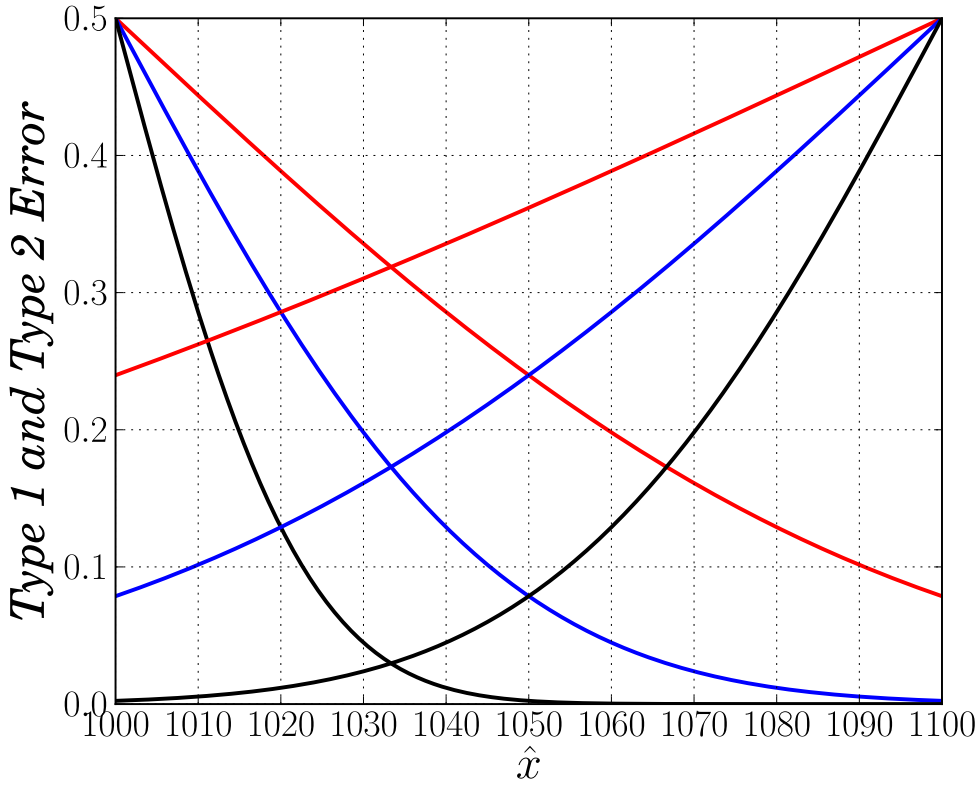
$(\mu_1, \sigma_1) = (1000, 100)$

Probability of Type 2 error vs. $\hat{x}$

$(\mu_2, \sigma_2) = (1100, 100)$ .

Sample sizes : 2 (*red*) , 8 (*blue*) , 32 (*black*) .



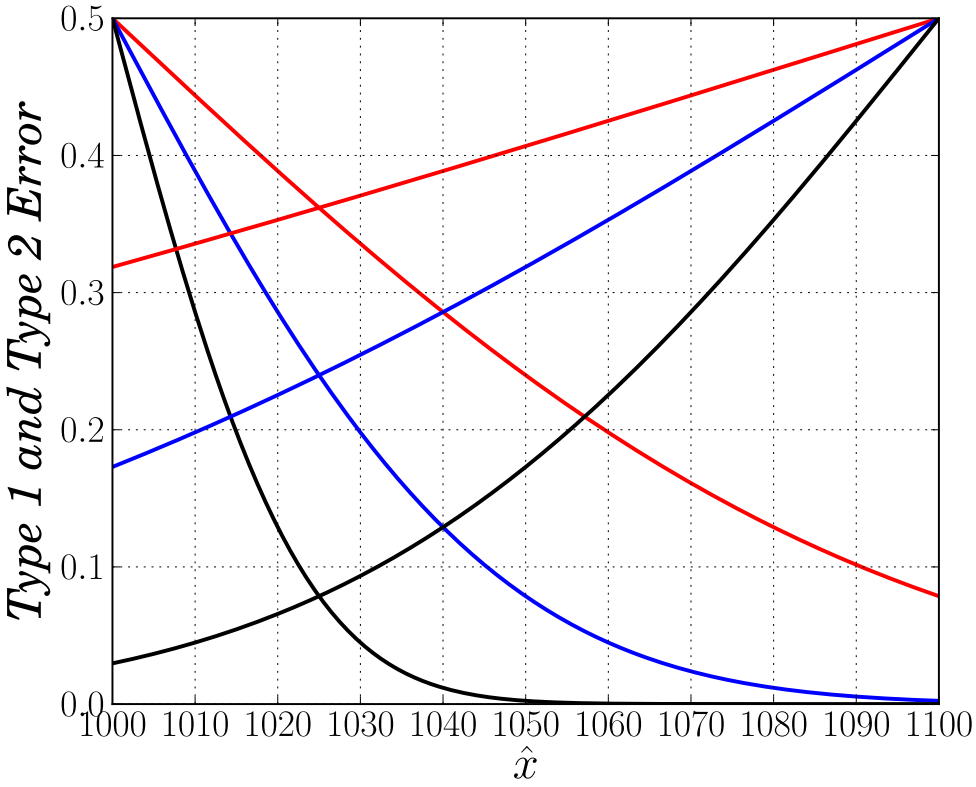

The probability of Type 1 and Type 2 errors versus $\hat{x}$ .

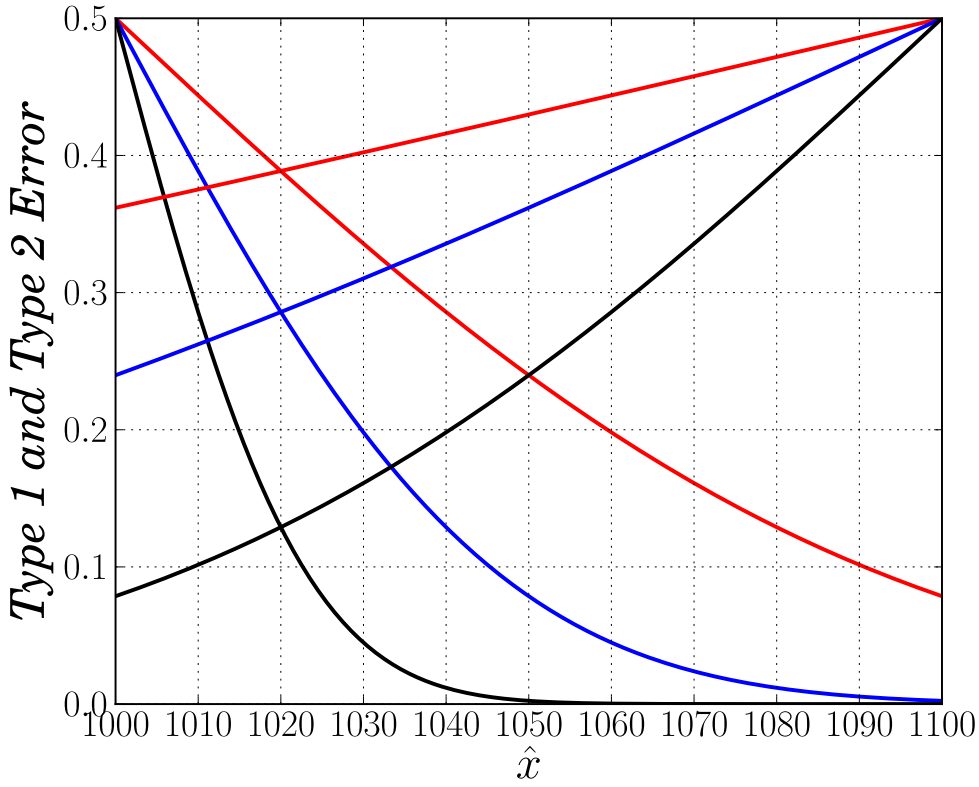
*Left* : $(\mu_1, \boldsymbol{\sigma}_1) = (1000, \mathbf{100})$, $(\mu_2, \boldsymbol{\sigma}_2) = (1100, \mathbf{100})$.

*Right* : $(\mu_1, \boldsymbol{\sigma}_1) = (1000, \mathbf{100})$, $(\mu_2, \boldsymbol{\sigma}_2) = (1100, \mathbf{200})$.

Colors indicate sample size : 2 (*red*), 8 (*blue*), 32 (*black*) .

Curves of a given color intersect at the *minimax* $\hat{x}$-value.




The probability of Type 1 and Type 2 errors versus $\hat{x}$ .

*Left* : $(\mu_1, \boldsymbol{\sigma}_1) = (1000, \mathbf{100})$, $(\mu_2, \boldsymbol{\sigma}_2) = (1100, \mathbf{300})$.

*Right* : $(\mu_1, \boldsymbol{\sigma}_1) = (1000, \mathbf{100})$, $(\mu_2, \boldsymbol{\sigma}_2) = (1100, \mathbf{400})$.

Colors indicate sample size : 2 (*red*), 8 (*blue*), 32 (*black*) .

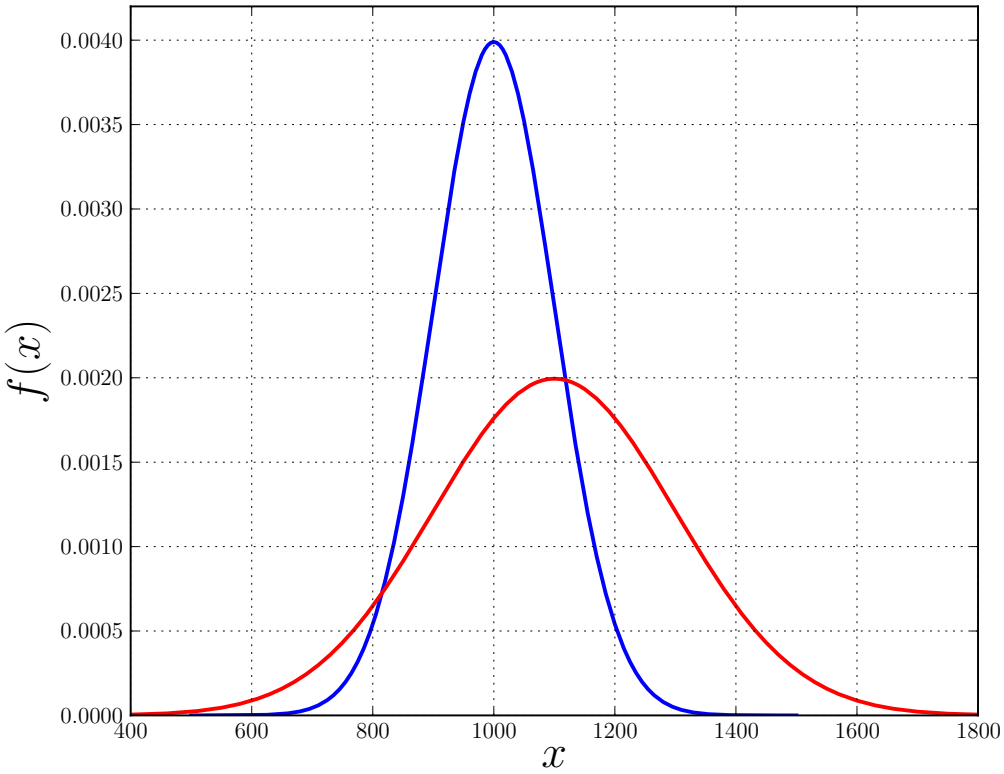
Curves of a given color intersect at the *minimax* $\hat{x}$-value.

**NOTE** :

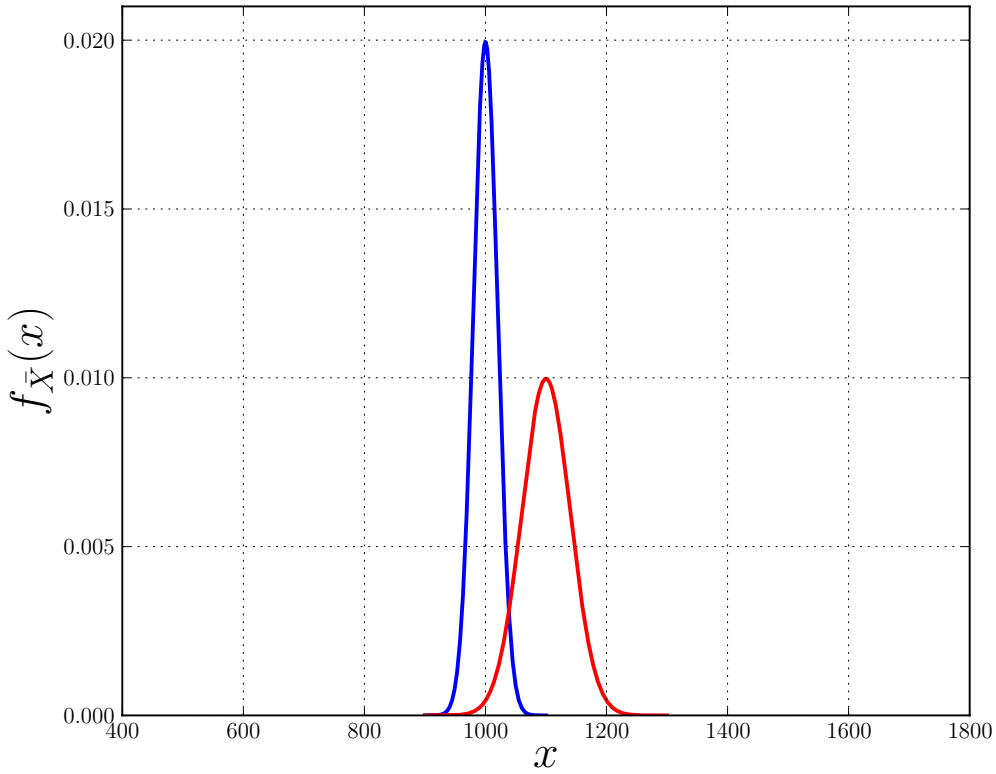
- There is an *optimal* value $\hat{x}^*$ of $\hat{x}$ .
- At $\hat{x}^*$ the value of

$$\max \{ P(\text{Type 1 Error}) , P(\text{Type 2 Error}) \}$$

  is *minimized* .
- We call $\hat{x}^*$ the *minimax* value.
- The value of $\hat{x}^*$ depends on $\sigma_1$ and $\sigma_2$ .
- The value of $\hat{x}^*$ is independent of the sample size.
- (We will prove this!)



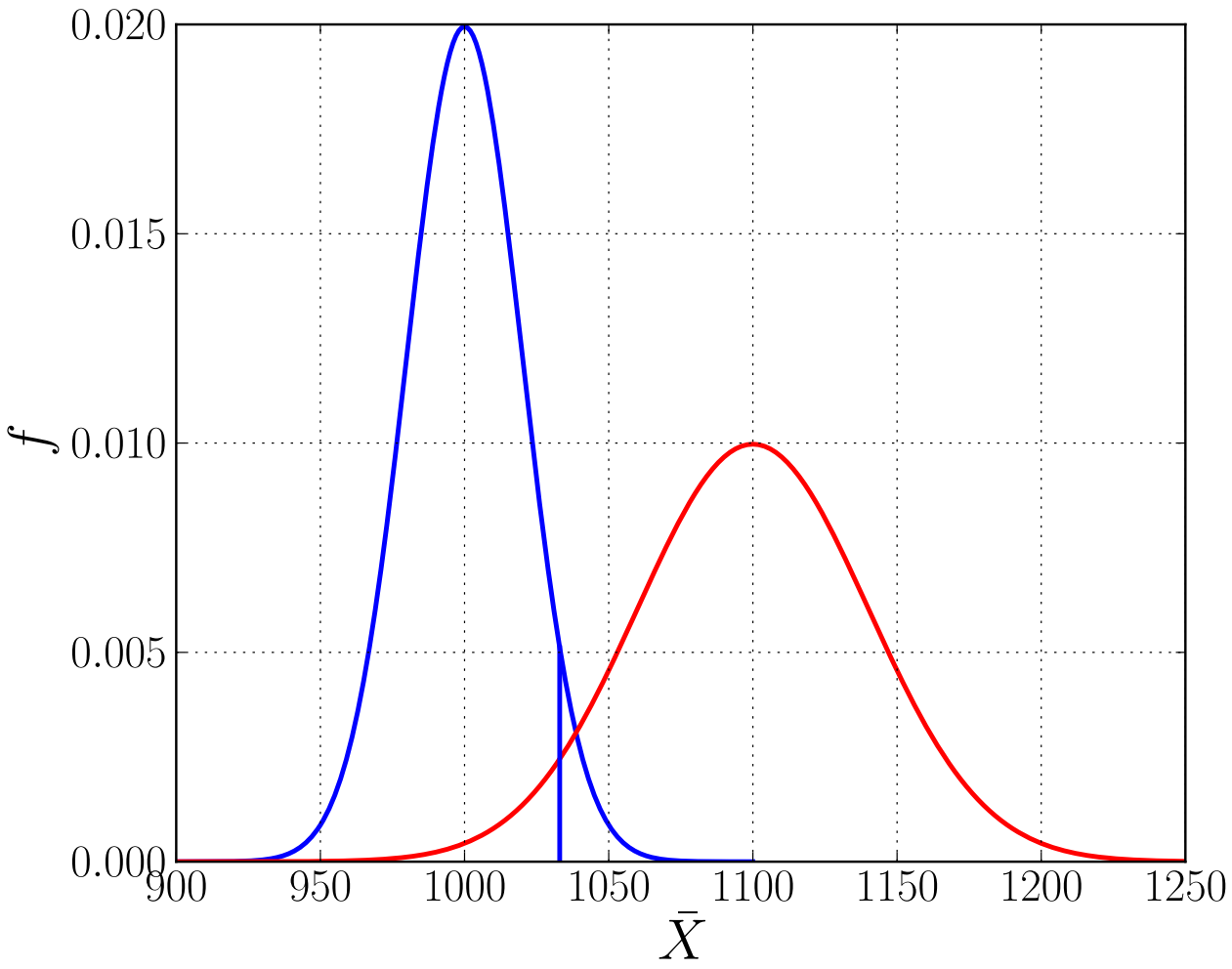

The population density functions.




The density functions of $\bar{X}$ $(n = 25)$ .

$$(\mu_1 \ , \ \sigma_1) \quad = \quad (1000 \ , \ 100) \qquad (\textit{blue})$$

$$(\mu_2 \ , \ \sigma_2) \quad = \quad (1100 \ , \ 200) \qquad (\textit{red})$$




The density functions of $\bar{X}$ $(n = 25)$ , with minimax value of $\hat{x}$ .

$(\mu_1, \sigma_1) = (1000, 100)$ $(blue)$ , $(\mu_2, \sigma_2) = (1100, 200)$ $(red)$ .

The minimax value $\hat{x}^*$ of $\hat{x}$ is *easily computed* : At $\hat{x}^*$ we have

$$P(\text{Type 1 Error}) \quad = \quad P(\text{Type 2 Error}) \ ,$$

$$\iff \qquad P(\bar{X} \geq \hat{x}^* \mid \mu = \mu_1) \quad = \quad P(\bar{X} \leq \hat{x}^* \mid \mu = \mu_2) \ ,$$

$$\iff \qquad \Phi\Big(\frac{\mu_1 - \hat{x}^*}{\sigma_1/\sqrt{n}}\Big) \quad = \quad \Phi\Big(\frac{\hat{x}^* - \mu_2}{\sigma_2/\sqrt{n}}\Big) \ ,$$

$$\iff \qquad \frac{\mu_1 - \hat{x}^*}{\sigma_1/\sqrt{n}} \quad = \quad \frac{\hat{x}^* - \mu_2}{\sigma_2/\sqrt{n}} \ , \quad (\text{by } \textit{monotonicity} \text{ of } \Phi \ ).$$

from which

$$\hat{x}^* \quad = \quad \frac{\mu_1 \cdot \sigma_2 \ + \ \mu_2 \cdot \sigma_1}{\sigma_1 + \sigma_2} \ . \qquad (\text{ Check ! })$$

With $\mu_1 = 1000$ , $\sigma_1 = 100$ , $\mu_2 = 1100$ , $\sigma_2 = 200$ , we have

$$\hat{x}^* \quad = \quad \frac{1000 \cdot 200 \ + \ 1100 \cdot 100}{100 + 200} \quad = \quad 1033 \ .$$

We have proved :

**THEOREM** : Suppose Lot 1 and Lot 2 are *normally distributed* , with mean and standard deviation

$$(\mu_1 \ , \ \sigma_1) \quad \text{and} \quad (\mu_2 \ , \ \sigma_2) \ , \quad \text{where} \quad (\mu_1 < \mu_2) \ ,$$

and *sample size* $n$ .

Then the value of *decision criterion* $\hat{x}$ that *minimizes*

$$\max \ \{ \ P(\text{Type 1 Error}) \ , \ P(\text{Type 2 Error}) \ \} \ ,$$

*i.e.*, the value of $\hat{x}$ that minimizes

$$\max \ \{ \ P(\bar{X} \geq \hat{x} \mid \mu = \mu_1, \ \sigma = \sigma_1) \ \ , \ \ P(\bar{X} \leq \hat{x} \mid \mu = \mu_2, \ \sigma = \sigma_2 \ \} \ ,$$

is given by

$$\hat{x}^* \quad = \quad \frac{\sigma_1 \mu_2 \ + \ \sigma_2 \mu_1}{\sigma_1 \ + \ \sigma_2} \ .$$

**EXERCISE** :

Determine the optimal *decision criterion* $\hat{x}^*$ that *minimizes*

$$\max \{ P(\text{Type 1 Error}) , P(\text{Type 2 Error}) \} ,$$

when

$$(\mu_1, \sigma_1) = (1000, 200) \quad , \quad (\mu_2, \sigma_2) = (1100, 300) .$$

For this $\hat{x}^*$ find the probability of a Type 1 and a Type 2 Error ,

when

$$n = 1 \quad , \quad n = 25 \quad , \quad n = 100 .$$

**EXAMPLE** ( *Known standard deviation* ) :

Suppose we have a sample of size $n = 9$ , with

$$\bar{X} \quad = \quad \mathbf{4.88} \ ,$$

from a *normal population* with standard deviation $\sigma = 0.2$ .

The *claim* is that the population mean equals

$$\mu = \mathbf{5.00} \ , \qquad ( \text{ the } \text{"}\textit{null hypothesis}\text{"} \ H_0 \ )$$

We see that $\quad | \bar{X} - \mu | \quad = \quad | 4.88 - 5.00 | \quad = \quad 0.12 \ .$

We *reject* $H_0$ if $P( \ | \bar{X} - \mu | \ \geq \ 0.12 \ )$ is *rather small* , say

$P( \ | \bar{X} - \mu | \ \geq \ 0.12 \ ) \ < \ 10 \ \% \qquad ( \text{"}\textit{level of significance}\text{"} \ \ 10 \ \% \ )$

Do we *accept* $H_0$ ?

**SOLUTION** ( *Known standard deviation* ) : We have

$$n = 9\,, \quad \sigma = 0.2\,, \quad \bar{X} = \mathbf{4.88}, \quad \mu = \mathbf{5.0}\,, \quad |\,\bar{X} - \mu\,| = 0.12\,.$$

Since

$$Z \equiv \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad \text{is standard normal ,}$$

the "*p-value*" ( from the *standard normal Table* ) is

$$P(\ |\,\bar{X} - \mu\,| \ \geq\ 0.12\ ) = P\Big(\ |\,Z\,| \ \geq\ \frac{0.12}{0.2/\sqrt{9}}\ \Big)$$

$$= P(\ |\,Z\,| \ \geq\ 1.8\ ) = 2\,\Phi(-1.8) \cong 7.18\ \%\ .$$

Thus we *reject* $H_0$ at level of significance 10 % .

**NOTE** : We would *accept* $H_0$ if the level of significance were 5 % .

( We are "more tolerant" when the level of significance is smaller. )

**EXAMPLE** ( *Unknown standard deviation, large sample* ) :

A sample of size $n = 64$ from a *normal population* has

$$\text{sample mean } \bar{X} = \mathbf{4.847} \ ,$$

and

$$\text{sample standard deviation } \hat{S} = 0.234 \ .$$

Test the *null hypothesis*

$$H_0 \ : \quad \mu \ \leq \ \mathbf{4.8} \ ,$$

and *reject* it if $P(\bar{X} \ \geq \ 4.847)$ is small, say if

$$P(\bar{X} \ \geq \ 4.847) \ < \ 5 \ \% \ .$$

**NOTE** : Since the sample size $n = 64$ is large, we can assume that

$$\sigma \ \cong \ \hat{S} \ = 0.234 \ .$$

**SOLUTION** ( *Unknown standard deviation, large sample* ) :

Given $\bar{X} = \mathbf{4.847}$ , $\mu = \mathbf{4.8}$ , $n = 64$ , $\sigma \cong \hat{S} = 0.234$ .

Using the standard normal approximation we have that

if and only if $$\bar{X} \geq 4.847 \, ,$$

$$Z \equiv \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{4.847 - 4.8}{0.234/8} = 1.6 \, .$$

From the standard normal Table we have the *p-value*

$$P(Z \geq 1.6) = 1 - \Phi(1.6) = \Phi(-1.6) = 5.48 \, \% \, .$$

**CONCLUSION**:

We (barely) *accept* $H_0$ at level of significance 5 % .

( We would *reject* $H_0$ at level of significance 10 % .)

**EXAMPLE** ( *Unknown standard deviation, small sample* ) :

A sample of size $n = 16$ from a *normal population* has

$$\text{sample mean } \bar{X} = \mathbf{4.88} \ ,$$

and

$$\text{sample standard deviation } \hat{S} = 0.234 \ .$$

Test the *null hypothesis*

$$H_0 \ : \quad \mu \ \leq \ \mathbf{4.8} \ ,$$

and *reject* it if

$$P(\bar{X} \ \geq \ \mathbf{4.88}) \ < \ 5 \ \% \ .$$

**NOTE** :

If $n \leq 30$ then the approximation $\sigma \ \cong \ \hat{S}$ is not so accurate.

In this case better use the "*student t-distribution*" $T_{n-1}$ .

**SOLUTION** ( *Unknown standard deviation, small sample* ) :

With $n = 16$ we have $\bar{X} \geq \mathbf{4.88}$ if and only if

$$T_{n-1} \;=\; T_{15} \;=\; \frac{\bar{X} - \mu}{\hat{S}/\sqrt{n}} \;\geq\; \frac{4.88 - 4.8}{0.234/4} \;\cong\; 1.37 \; .$$

A *t-distribution Table* shows that

$$P(\, T_{15} \geq 1.341 \,) \;=\; 10 \; \% \; ,$$

$$P(\, T_{15} \geq 1.753 \,) \;=\; 5 \; \% \; .$$

Thus we *reject* $H_0$ at *level of significance* 10 % ,

but we *accept* $H_0$ at *level of significance* 5 % .

(We are "more tolerant" when the level of significance is smaller.)

**EXAMPLE** ( *Testing a hypothesis on the standard deviation* ) :

A sample of 16 items has sample standard deviation $\hat{S} = \mathbf{2.58}$ .

Do you believe the population standard deviation satisfies $\sigma \leq \mathbf{2.0}$ ?

**SOLUTION** : We already know that

$$\frac{n-1}{\sigma^2}\,\hat{S}^2 \quad \text{has the} \quad \chi^2_{n-1} \quad \text{distribution .}$$

For our data

$$\hat{S} \;\geq\; 2.58 \qquad \text{if and only if} \qquad \frac{n-1}{\sigma^2}\,\hat{S}^2 \;\geq\; \frac{15}{4}\,2.58^2 \;=\; 24.96 \;,$$

and from the $\chi^2$ *Table*

$$P(\chi^2_{15} \;\geq\; 25.0) \quad \cong \quad 5.0\ \% \;.$$

Thus we (barely) *accept* the hypothesis at significance level 5 % .

We would *reject* the hypothesis at significance level 10 % .

**EXERCISE** :

A sample of 16 items has sample standard deviation

$$\hat{S} \quad = \quad 0.83 \ .$$

Do you believe the *hypothesis* that $\sigma$ satisfies

$$\sigma \quad \leq \quad 1.2 \quad ?$$

( Probably Yes ! )

# LEAST SQUARES APPROXIMATION

## Linear Least Squares

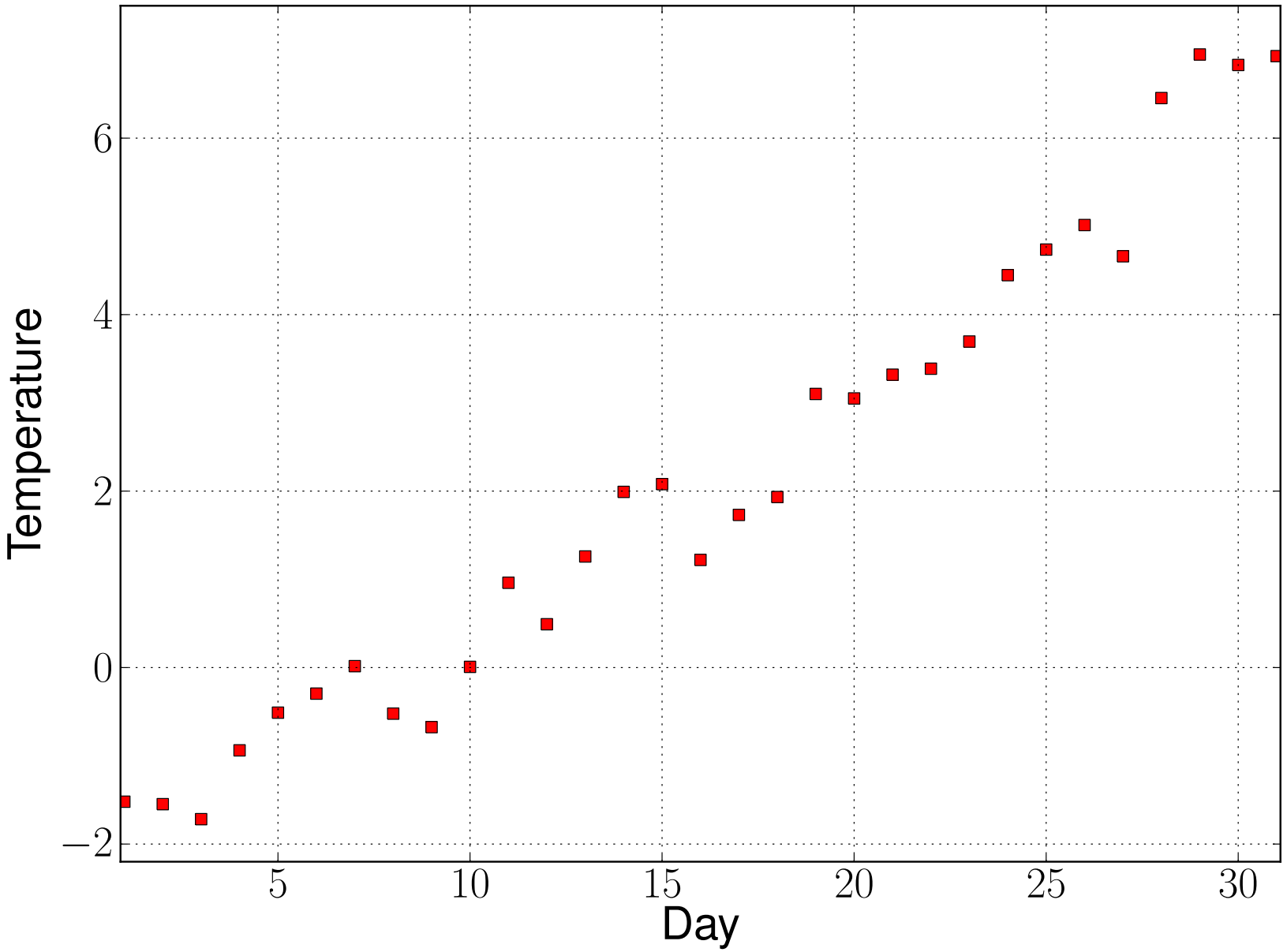
Recall the following data :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | -1.52 | 8 | -0.52 | 15 | 2.08 | 22 | 3.39 | 29 | 6.95 |
| 2 | -1.55 | 9 | -0.67 | 16 | 1.22 | 23 | 3.69 | 30 | 6.83 |
| 3 | -1.72 | 10 | 0.01 | 17 | 1.73 | 24 | 4.45 | 31 | 6.93 |
| 4 | -0.94 | 11 | 0.96 | 18 | 1.93 | 25 | 4.74 | | |
| 5 | -0.51 | 12 | 0.49 | 19 | 3.10 | 26 | 5.01 | | |
| 6 | -0.29 | 13 | 1.26 | 20 | 3.05 | 27 | 4.66 | | |
| 7 | 0.02 | 14 | 1.99 | 21 | 3.32 | 28 | 6.45 | | |

Average daily high temperature in Montreal in March : 1943-2014 .

( Source : `http://climate.weather.gc.ca/` )

These data have *sample correlation coefficient* $R_{X,Y} = \mathbf{0.98}$ .



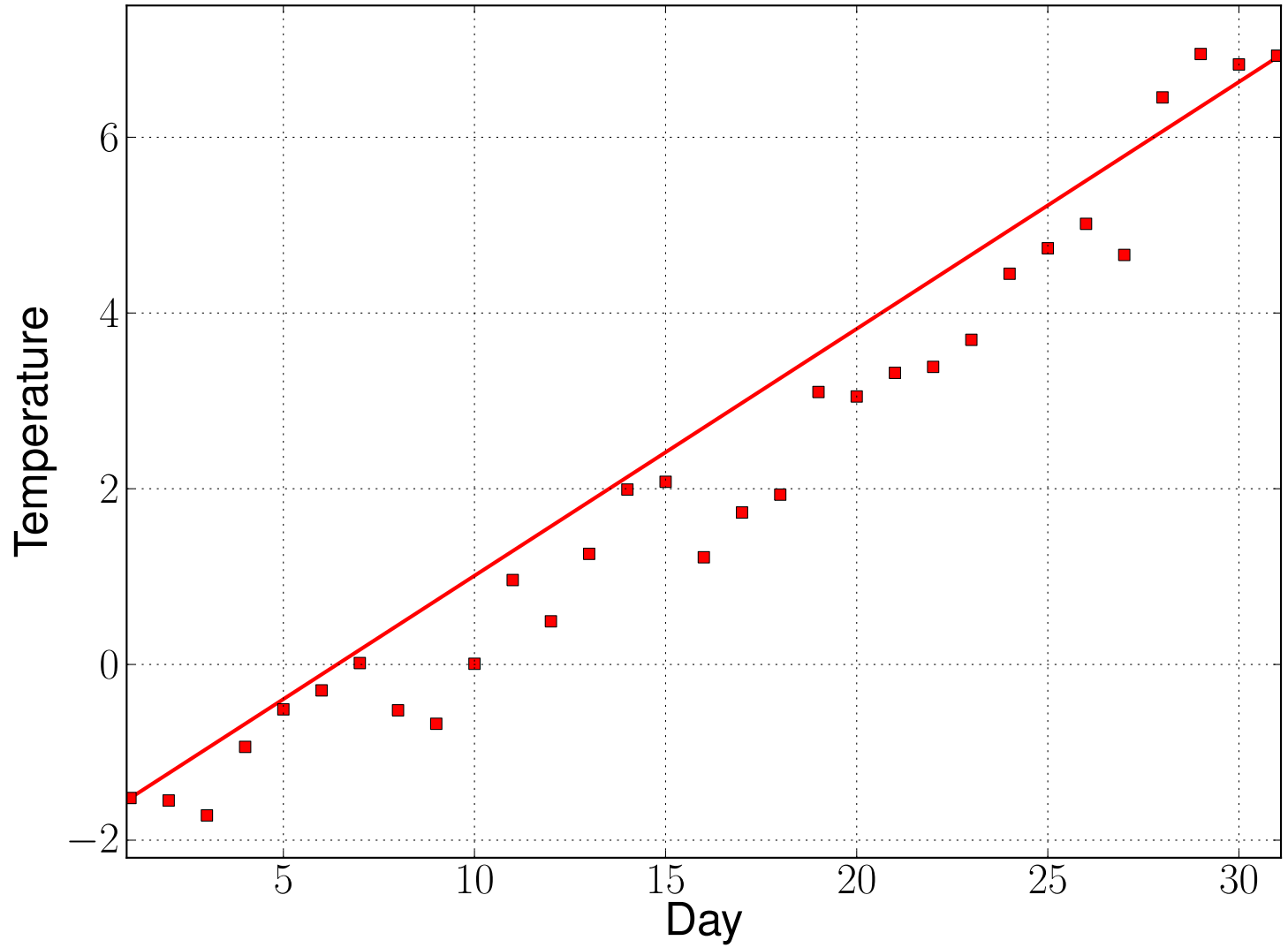

Average daily high temperature in Montreal in March

Suppose that :

- We believe that these temperatures basically increase *linearly* .
- In fact we found the *sample correlation coefficient* $R_{xy} = 0.98$ .
- Thus we believe in a relation

$$T_k \quad = \quad c_1 \quad + \quad c_2 \, k \; , \qquad k \; = \; 1, \; 2, \; \cdots \; , 31 \; .$$

- The *deviations* from linearity come from *random influences* .
- These random influences can be due to *many factors* .
- The deviations may have a *normal distribution* .
- We want to determine "*the best* " linear approximation.




Average daily high temperatures, with a *linear approximation* .

**QUESTION** : Guess how this linear approximation was obtained !

- There are many ways to determine such a linear approximation.

- Often used is the *least squares method* .

- This method determines $c_1$ and $c_2$ that *minimize*

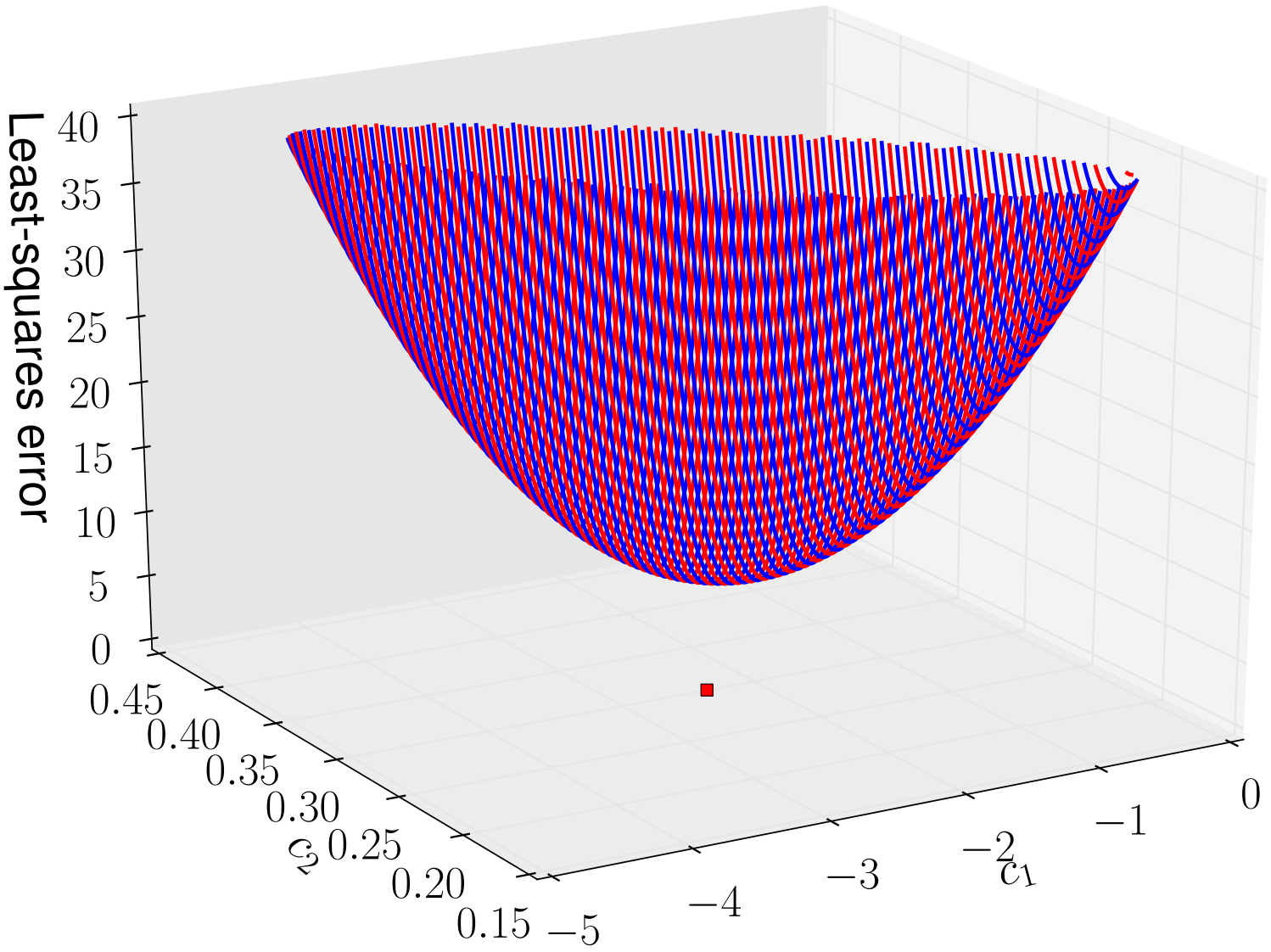
$$\sum_{k=1}^{N} ( \, T_k \, - \, (c_1 \, + \, c_2 x_k) \, )^2 \, ,$$

  where, in our example, $N = 31$ and $x_k = k$ .

- To do so set the *partial derivatives* w.r.t. $c_1$ and $c_2$ to zero :

$$\text{w.r.t. } c_1 \; : \qquad -\, 2 \sum_{k=1}^{N} ( \, T_k \, - \, (c_1 \, + \, c_2 x_k) \, ) \; = \; 0 \, ,$$

$$\text{w.r.t. } c_2 \; : \qquad -\, 2 \sum_{k=1}^{N} x_k \, ( \, T_k \, - \, (c_1 \, + \, c_2 x_k) \, ) \; = \; 0 \, .$$




The least squares error versus $c_1$ and $c_2$ .

From setting the partial derivatives to zero, we have

$$\sum_{k=1}^{N}( T_k - (c_1 + c_2 x_k) ) = 0 \quad , \quad \sum_{k=1}^{N} x_k ( T_k - (c_1 + c_2 x_k) ) = 0 .$$

Solving these two equations for $c_1$ and $c_2$ gives

$$c_2 = \frac{\sum_{k=1}^{N} x_k T_k - \bar{x} \sum_{k=1}^{N} T_k}{\sum_{k=1}^{N} x_k^2 - N \bar{x}^2} ,$$

and

$$c_1 = \bar{T} - c_2 \, \bar{x} ,$$

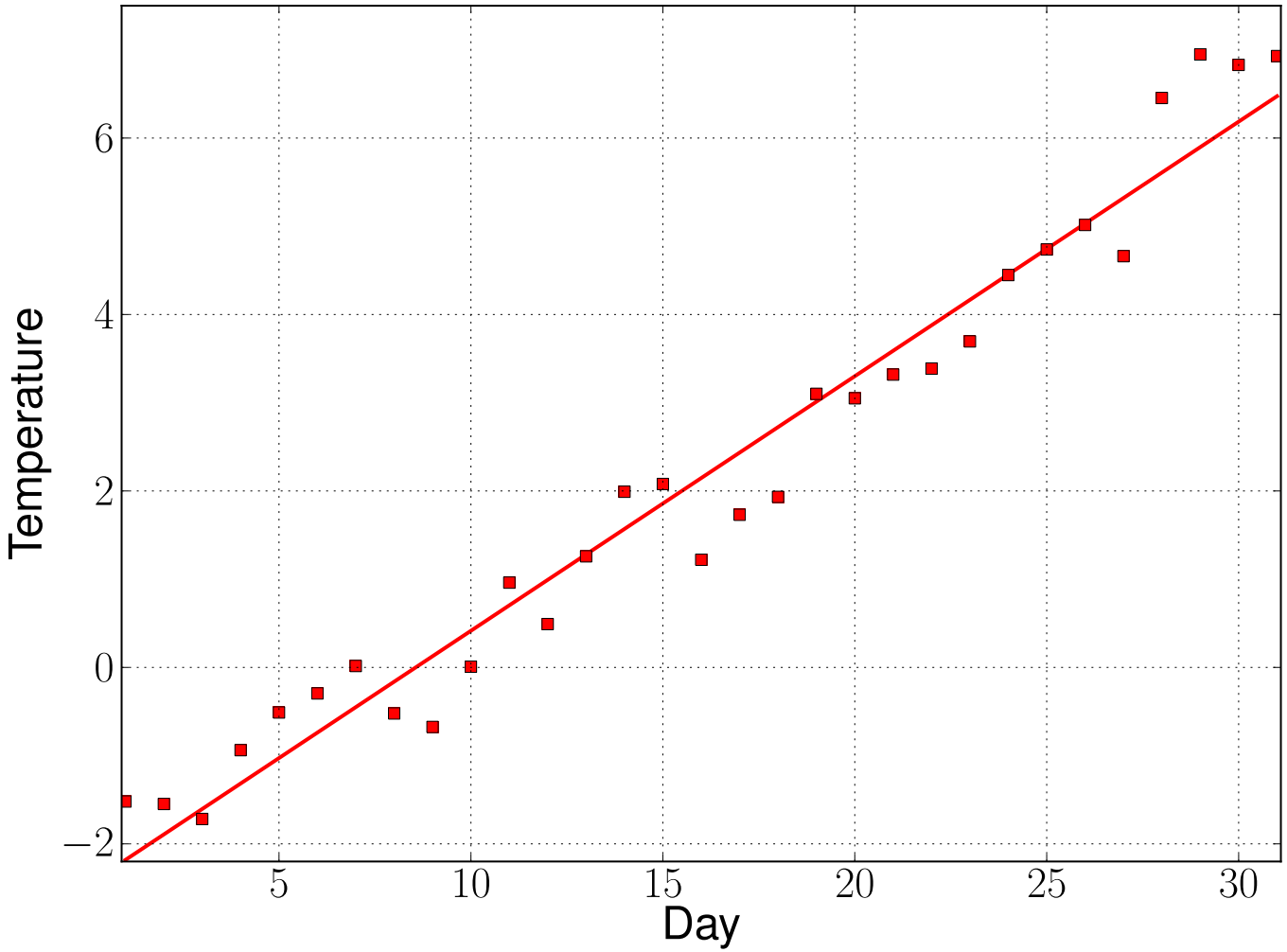
where

$$\bar{x} = \frac{1}{N} \sum_{k=1}^{N} x_k \quad , \quad \bar{T} = \frac{1}{N} \sum_{k=1}^{N} T_k .$$

**EXERCISE** : Check these formulas !

**EXAMPLE** : For our "*March temperatures*" example, we find

$$c_1 = -2.47 \quad \text{and} \quad c_2 = 0.289 \ .$$




Average daily high temperatures, with linear *least squares approximation* .

## General Least Squares

Given discrete data points

$$\{ (x_i, y_i) \}_{i=1}^{N} ,$$

find the coefficients $c_k$ of the function

$$p(x) \equiv \sum_{k=1}^{n} c_k \, \phi_k(x) ,$$

that *minimize* the *least squares error*

$$E_L \equiv \sum_{i=1}^{N} \left( p(x_i) - y_i \right)^2$$

**EXAMPLES** :

- $p(x) = c_1 + c_2 \, x .$ (Already done !)
- $p(x) = c_1 + c_2 \, x + c_3 \, x^2 .$ (Quadratic approximation)

For any vector $\mathbf{x} \in \mathbb{R}^N$ let

$$\| \mathbf{x} \|^2 \quad \equiv \quad \mathbf{x}^T\mathbf{x} \quad \equiv \quad \sum_{i=1}^{N} x_k^2 \ . \qquad (T \text{ denotes } \textit{transpose}).$$

Then

$$E_L \ \equiv \ \sum_{i=1}^{N} [\ p(x_i) - y_i\ ]^2 \ = \ \| \begin{pmatrix} p(x_1) \\ \cdot \\ \cdot \\ p(x_N) \end{pmatrix} - \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_N \end{pmatrix} \|^2$$

$$= \ \| \begin{pmatrix} \sum_{i=1}^{n} c_i\phi_i(x_1) \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} c_i\phi_i(x_N) \end{pmatrix} - \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_N \end{pmatrix} \|^2$$

$$= \ \| \begin{pmatrix} \phi_1(x_1) & \cdot & \phi_n(x_1) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \phi_1(x_N) & \cdot & \phi_n(x_N) \end{pmatrix} \begin{pmatrix} c_1 \\ \cdot \\ c_n \end{pmatrix} - \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_N \end{pmatrix} \|^2 \quad \equiv \quad \| \mathbf{Ac} - \mathbf{y} \|^2 \ .$$

**THEOREM** :

For the least squares error $E_L$ to be *minimized* we must have

$$\mathbf{A}^T \mathbf{A}\, \mathbf{c} = \mathbf{A}^T \mathbf{y} .$$

**PROOF** :

$$\begin{aligned} E_L &= \| \mathbf{Ac} - \mathbf{y} \|^2 \\ &= (\mathbf{Ac} - \mathbf{y})^T (\mathbf{Ac} - \mathbf{y}) \\ &= (\mathbf{Ac})^T \mathbf{Ac} - (\mathbf{Ac})^T \mathbf{y} - \mathbf{y}^T \mathbf{Ac} + \mathbf{y}^T \mathbf{y} \\ &= \mathbf{c}^T \mathbf{A}^T \mathbf{Ac} - \mathbf{c}^T \mathbf{A}^T \mathbf{y} - \mathbf{y}^T \mathbf{Ac} + \mathbf{y}^T \mathbf{y} . \end{aligned}$$

**PROOF** : ( continued $\cdots$ )

We had

$$E_L \;=\; \mathbf{c}^T\mathbf{A}^T\mathbf{A}\mathbf{c} \;-\; \mathbf{c}^T\mathbf{A}^T\mathbf{y} \;-\; \mathbf{y}^T\mathbf{A}\mathbf{c} \;+\; \mathbf{y}^T\mathbf{y}\;.$$

For a *minimum* we need

$$\frac{\partial E_L}{\partial \mathbf{c}} \;=\; 0, \qquad i.e., \qquad \frac{\partial E_L}{\partial c_i} \;=\; 0\;, \qquad i \;=\; 0, 1, \cdots, n\;,$$

which gives

$$\mathbf{c}^T\mathbf{A}^T\mathbf{A} \;+\; (\mathbf{A}^T\mathbf{A}\mathbf{c})^T \;-\; (\mathbf{A}^T\mathbf{y})^T \;-\; \mathbf{y}^T\mathbf{A} \;=\; 0\;, \quad (\text{ Check ! })$$

*i.e.*,

$$2\mathbf{c}^T\mathbf{A}^T\mathbf{A} \;-\; 2\mathbf{y}^T\mathbf{A} \;=\; 0\;,$$

or

$$\mathbf{c}^T\mathbf{A}^T\mathbf{A} \;=\; \mathbf{y}^T\mathbf{A}\;.$$

Transposing gives

$$\mathbf{A}^T\mathbf{A}\mathbf{c} \;=\; \mathbf{A}^T\mathbf{y}\;. \qquad\qquad \textbf{QED !}$$

**EXAMPLE** : Given the data points

$$\{ (x_i, y_i) \}_{i=1}^{4} \;=\; \{ (0,1) \,,\; (1,3) \,,\; (2,2) \,,\; (4,3) \} \,,$$

find the coefficients $c_1$ and $c_2$ of $p(x) = c_1 + c_2 x$ ,
that minimize

$$E_L \;\equiv\; \sum_{i=1}^{4} [ \, (c_1 + c_2 x_i) \;-\; y_i ]^2 \,.$$

**SOLUTION** : Here $N = 4$ , $n = 2$ , $\phi_1(x) = 1$ , $\phi_2(x) = x$ .

Use the Theorem :

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \;=\; \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 2 \\ 3 \end{pmatrix} ,$$

or

$$\begin{pmatrix} 4 & 7 \\ 7 & 21 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \;=\; \begin{pmatrix} 9 \\ 19 \end{pmatrix} ,$$

with solution $c_1 = 1.6$ and $c_2 = 0.371429$ .

**EXAMPLE** : Given the same data points, find the coefficients of

$$p(x) \;=\; c_1 \;+\; c_2 x \;+\; c_3 x^2 \;,$$

that minimize

$$E_L \;\equiv\; \sum_{i=1}^{4} \left[\, (c_1 \;+\; c_2\, x_i \;+\; c_3\, x_i^2) \;-\; y_i \,\right]^2 .$$

**SOLUTION** : Here

$$N = 4 \quad , \quad n = 3 \quad , \quad \phi_1(x) = 1 \quad , \quad \phi_2(x) = x \quad , \quad \phi_3(x) = x^2 \;.$$

Use the Theorem :

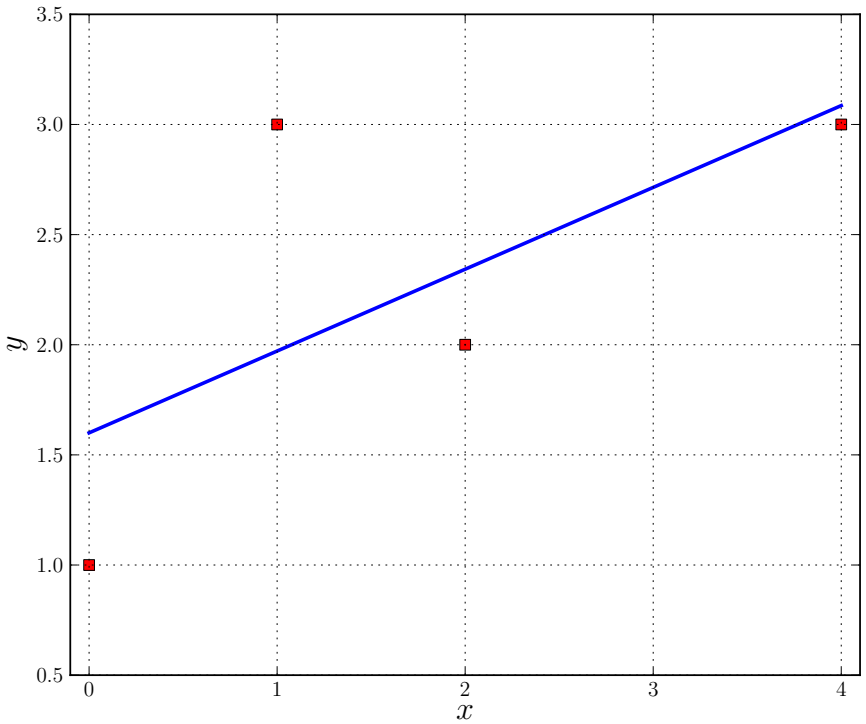
$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 4 \\ 0 & 1 & 4 & 16 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 4 & 16 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \;=\; \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 4 \\ 0 & 1 & 4 & 16 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 2 \\ 3 \end{pmatrix} ,$$

or

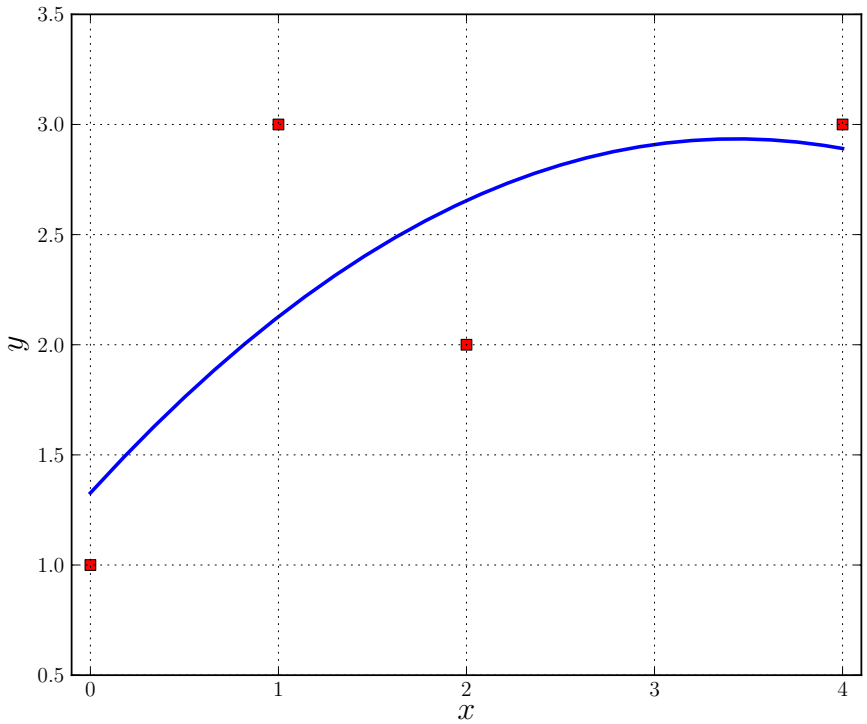
$$\begin{pmatrix} 4 & 7 & 21 \\ 7 & 21 & 73 \\ 21 & 73 & 273 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \;=\; \begin{pmatrix} 9 \\ 19 \\ 59 \end{pmatrix} ,$$

with solution $c_1 = 1.32727$ , $c_2 = 0.936364$ , $c_3 = -0.136364$ .

The least squares approximations from the preceding two examples :




$$p(x) = c_1 + c_2 x$$



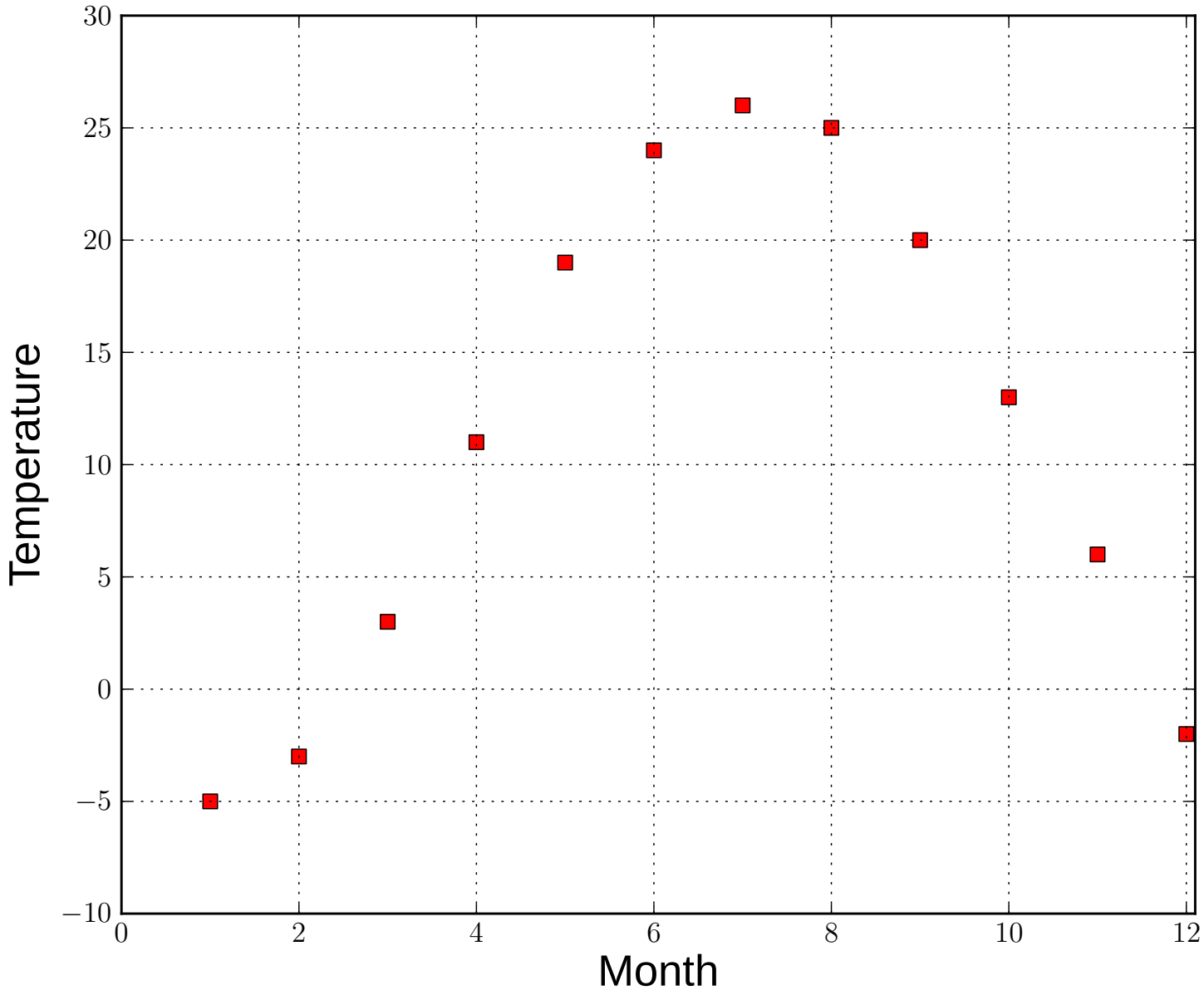

$$p(x) = c_1 + c_2 x + c_3 x^2$$

**EXAMPLE** : From actual data :

The average daily high temperatures in Montreal (by month) are :

| | |
|---|---|
| January | -5 |
| February | -3 |
| March | 3 |
| April | 11 |
| May | 19 |
| June | 24 |
| July | 26 |
| August | 25 |
| September | 20 |
| October | 13 |
| November | 6 |
| December | -2 |

Source : `http : //weather.uk.msn.com`




Average daily high temperature in Montreal (by month).

**EXAMPLE** : ( continued $\cdots$ )

The graph suggests using a 3-term *least squares approximation*

$$p(x) \quad = \quad c_1\ \phi_1(x) \quad + \quad c_2\ \phi_2(x) \quad + \quad c_3\ \phi_3(x)\ ,$$

of the form

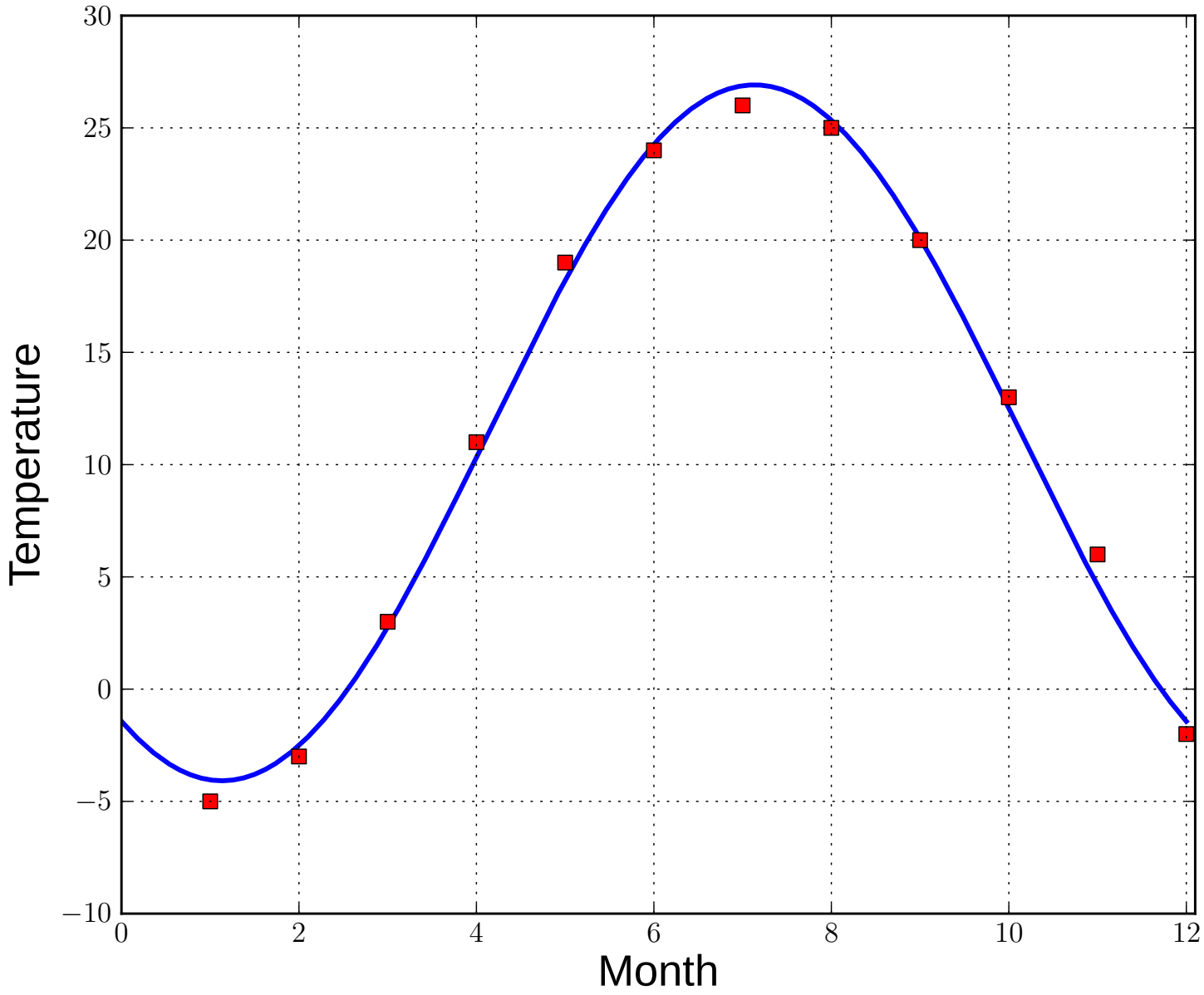
$$p(x) \quad = \quad c_1 \quad + \quad c_2\ \sin(\frac{\pi x}{6}) \quad + \quad c_3\ \cos(\frac{\pi x}{6})\ .$$

**QUESTIONS** :

- Why include $\phi_2(x) \ = \ \sin(\frac{\pi x}{6})$ ?
- Why is the argument $\frac{\pi x}{6}$ ?
- Why include the constant term $\phi_1(x) \ = \ c_1$ ?
- Why include $\phi_3(x) \ = \ \cos(\frac{\pi x}{6})$ ?

In this example we find the least squares coefficients

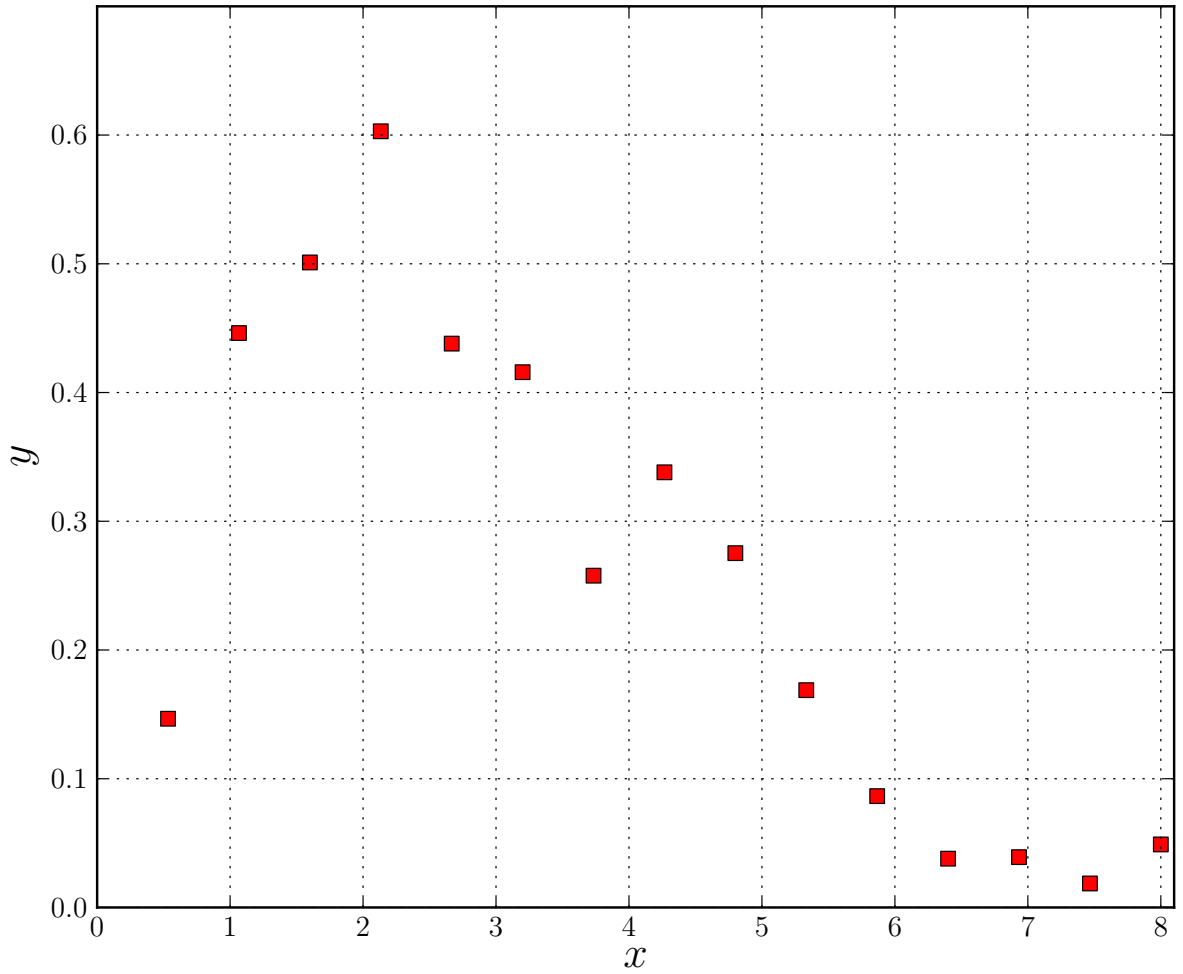
$$c_1 = 11.4 \quad , \quad c_2 = -8.66 \quad , \quad c_3 = -12.8\ .$$




Least squares fit of average daily high temperatures.

**EXAMPLE** :

Consider the following *experimental data* :



**EXAMPLE** : ( continued $\cdots$ )

Suppose we are given that :

- These data contain "*noise*" .
- The underlying physical process is understood.
- The *functional dependence* is *known* to have the form

$$y \quad = \quad c_1 \; x^{c_2} \; e^{-c_3 x} \; .$$

- The values of $c_1$ , $c_2$ , $c_3$ are *not* known.

**EXAMPLE** : ( continued $\cdots$ )

The functional relationship has the form

$$y \quad = \quad c_1 \; x^{c_2} \; e^{-c_3 x} \; .$$

**NOTE** :

- The unknown coefficients $c_1$ , $c_2$ , $c_3$ appear *nonlinearly* !
- This gives *nonlinear equations* for $c_1$ , $c_2$ , $c_3$ !
- Such problems are more *difficult* to solve !
- What to do ?

**EXAMPLE** : ( continued $\cdots$ )

Fortunately, in this example we can take the *logarithm* :

$$\log y \;=\; \log\,(\; c_1\; x^{c_2}\; e^{-c_3 x}\;) \;=\; \log c_1 \;+\; c_2 \log x \;-\; c_3\; x\;.$$

This gives a *linear* relationship

$$\log y \;=\; \hat{c}_1\;\phi_1(x) \;+\; c_2\;\phi_2(x) \;+\; c_3\;\phi_3(x)\;,$$

where

$$\hat{c}_1 \;=\; \log c_1\;.$$

and

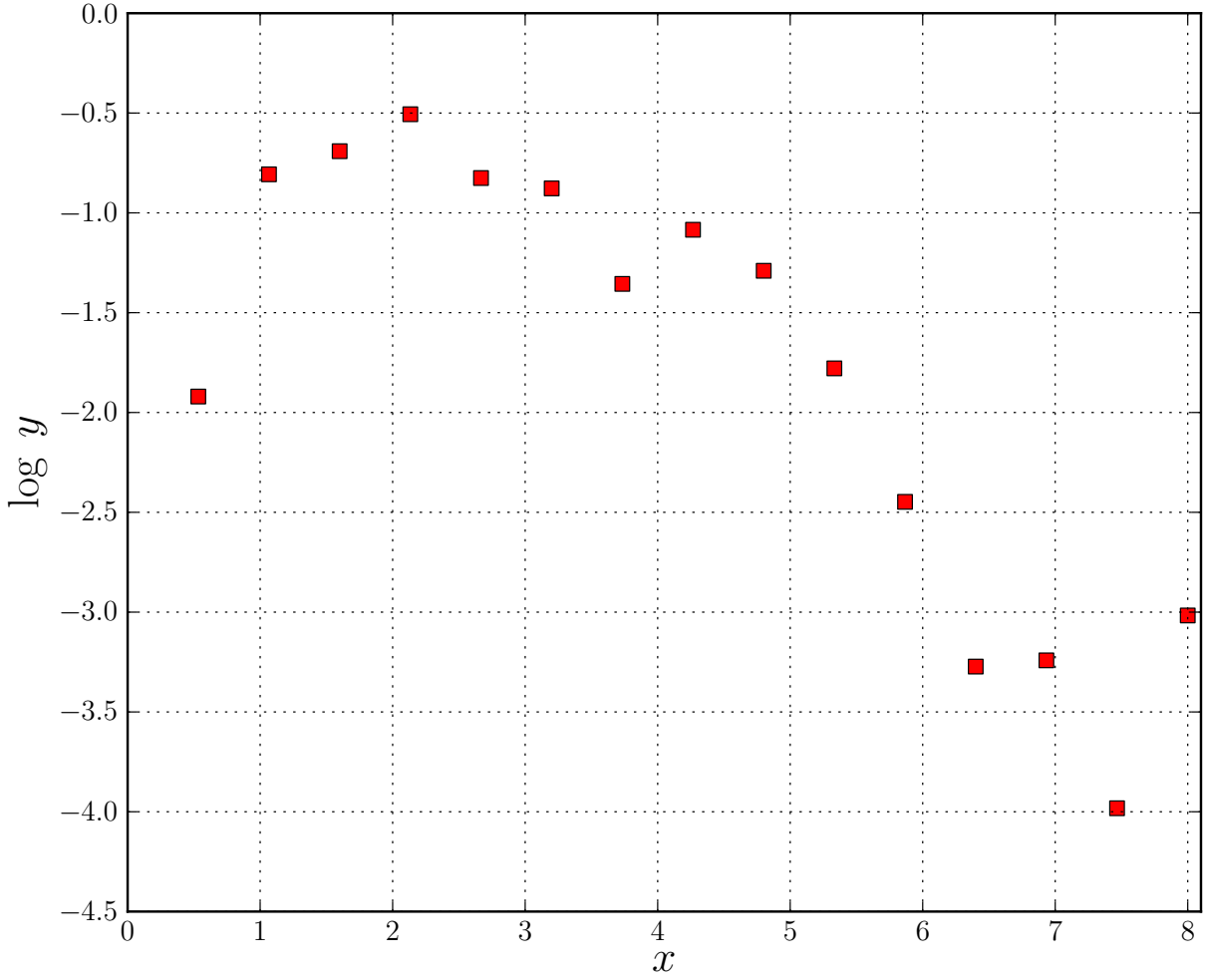
$$\phi_1(x) \;=\; 1 \quad,\quad \phi_2(x) \;=\; \log x \quad,\quad \phi_3(x) \;=\; -\,x\;.$$

Thus

- We can now use regular least squares.

- We first need to take the logarithm of the data.

## EXAMPLE : ( continued $\cdots$ )




The logarithm of the original $y$-values versus $x$ .

**EXAMPLE** : ( continued $\cdots$ )

We had

$$y \quad = \quad c_1 \; x^{c_2} \; e^{-c_3 x} \; ,$$

and

$$\log y \quad = \quad \hat{c}_1 \; \phi_1(x) \quad + \quad c_2 \; \phi_2(x) \quad + \quad c_3 \; \phi_3(x) \; ,$$

with

$$\phi_1(x) \; = \; 1 \quad , \quad \phi_2(x) \; = \; \log x \quad , \quad \phi_3(x) \; = \; - \, x \; ,$$

and

$$\hat{c}_1 \quad = \quad \log c_1 \; .$$

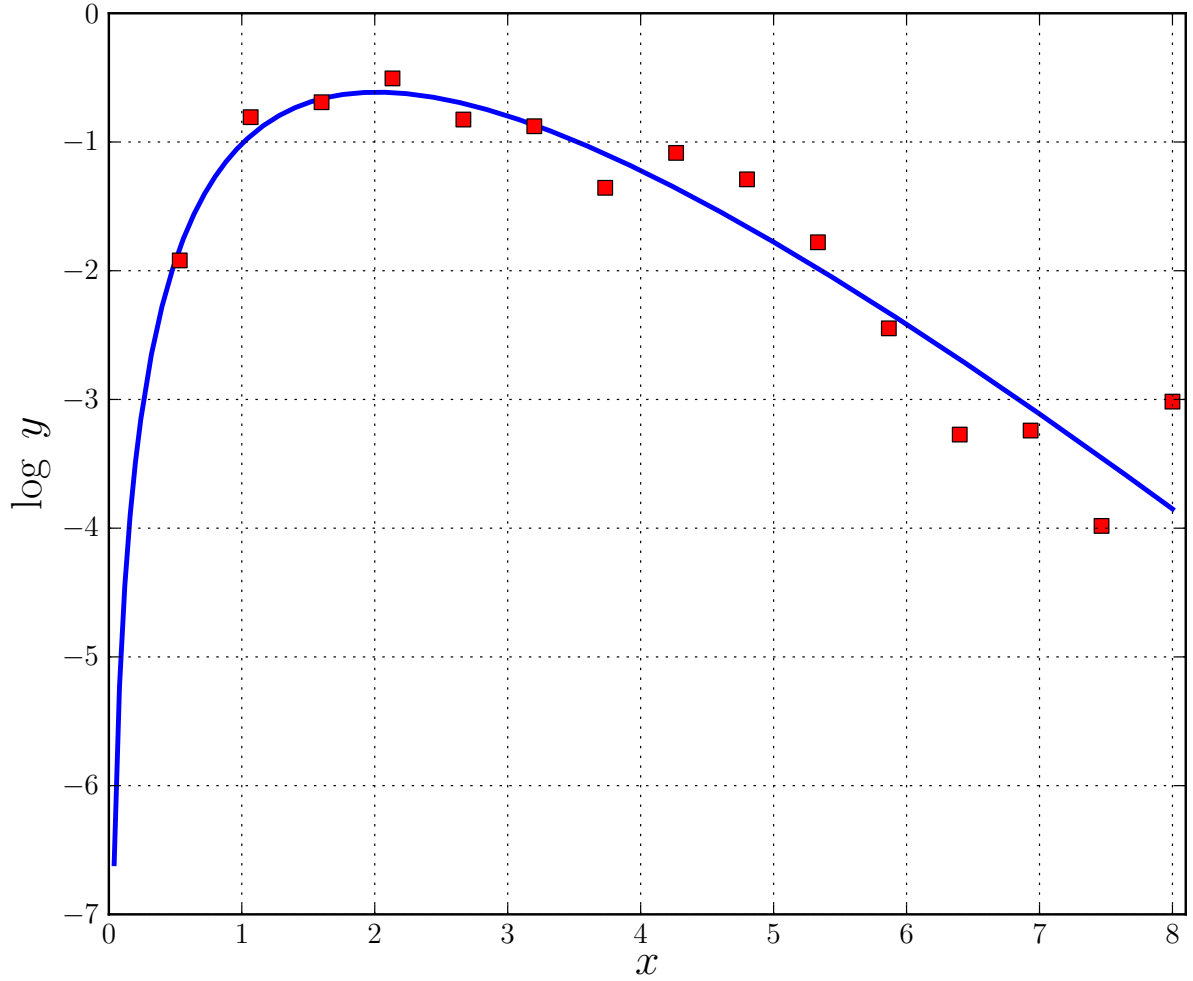
We find the following least squares values of the coefficients :

$$\hat{c}_1 \; = \; - \, 0.00473 \quad , \quad c_2 \; = \; 2.04 \quad , \quad c_3 \; = \; 1.01 \; ,$$

and

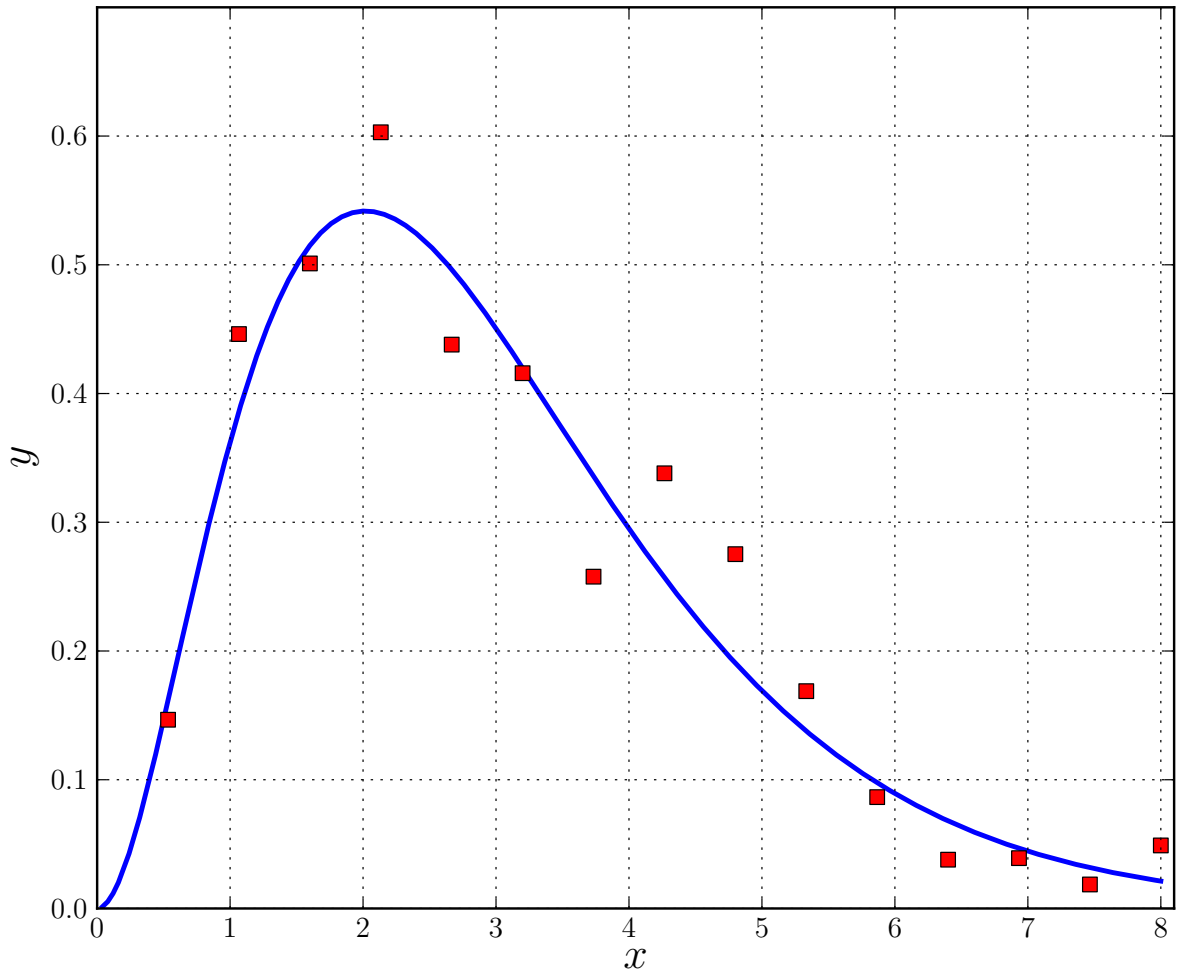
$$c_1 \quad = \quad e^{\hat{c}_1} \quad = \quad 0.995 \; .$$

## EXAMPLE : ( continued $\cdots$ )




The least squares approximation of the transformed data.

## EXAMPLE : ( continued $\cdots$ )




The least squares approximation shown in the original data.

# RANDOM NUMBER GENERATION

- Measured data often have *random fluctuations* .
- This may be due to *inaccurate measurements* .
- It may also be due to other *external influences* .

- Often we know or believe there is a *deterministic model* .
- (*i.e.*, the process can be modeled by a deterministic *equation* .)

- However, deterministic equations can also have *random behavior* !

- The study of such equations is sometimes called *chaos theory* .
- We will look at a simple example, namely, the *logistic equation* .

## The Logistic Equation

A simple deterministic model of *population growth* is

$$x_{k+1} \quad = \quad \lambda \, x_k \ , \qquad k = 1, 2, \ \cdots \ ,$$

for given $\lambda$ , $(\lambda \geq 0)$ , and for given $x_0$ , $(x_0 \geq 0)$ .

The solution is

$$x_k \quad = \quad \lambda^k \ x_0 \ , \qquad k = 1, 2, \ \cdots \ .$$

Thus

If $0 \leq \lambda < 1$ then $x_k \to 0$ as $k \to \infty$ ("extinction") .

If $\lambda > 1$ then $x_k \to \infty$ as $k \to \infty$ ("exponential growth") .

A more realistic population growth model is

$$x_{k+1} \quad = \quad \lambda \, x_k \, (1 - x_k) \, , \qquad k = 1, 2, \; \cdots \; ,$$

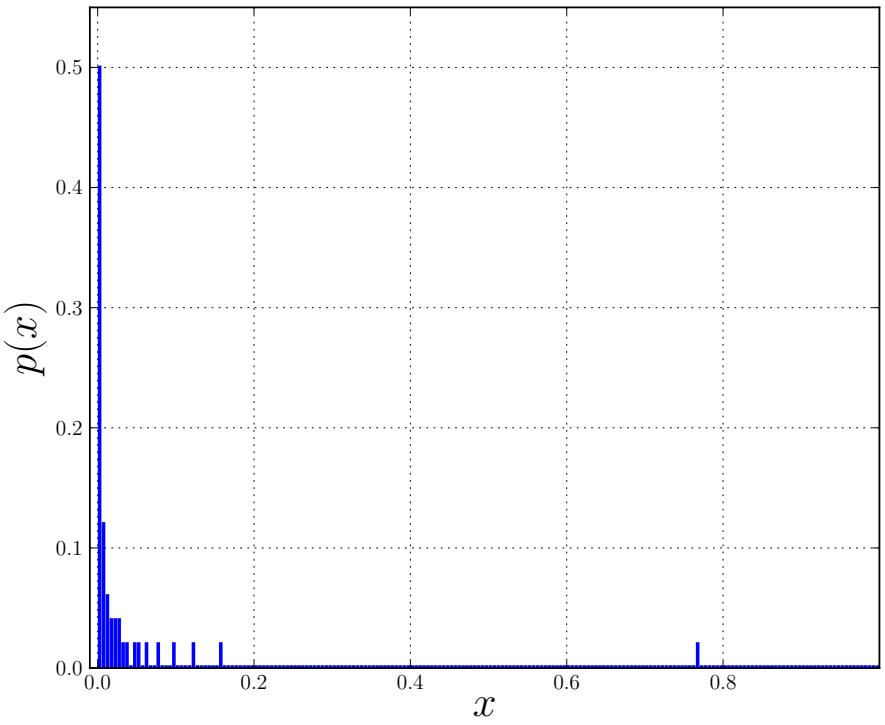
- This model is known as the *logistic equation* .
- The *maximum sustainable population* is $1$ .
- $\lambda$ is given, $(0 \leq \lambda \leq 4)$ .
- $x_0$ is given, $(0 \leq x_0 \leq 1)$ .
- Then $0 \leq x_k \leq 1$ for all $k$ . (Prove this !)

**QUESTION** : How does the sequence $\{x_k\}_{k=1}^{\infty}$ depend on $\lambda$ ?

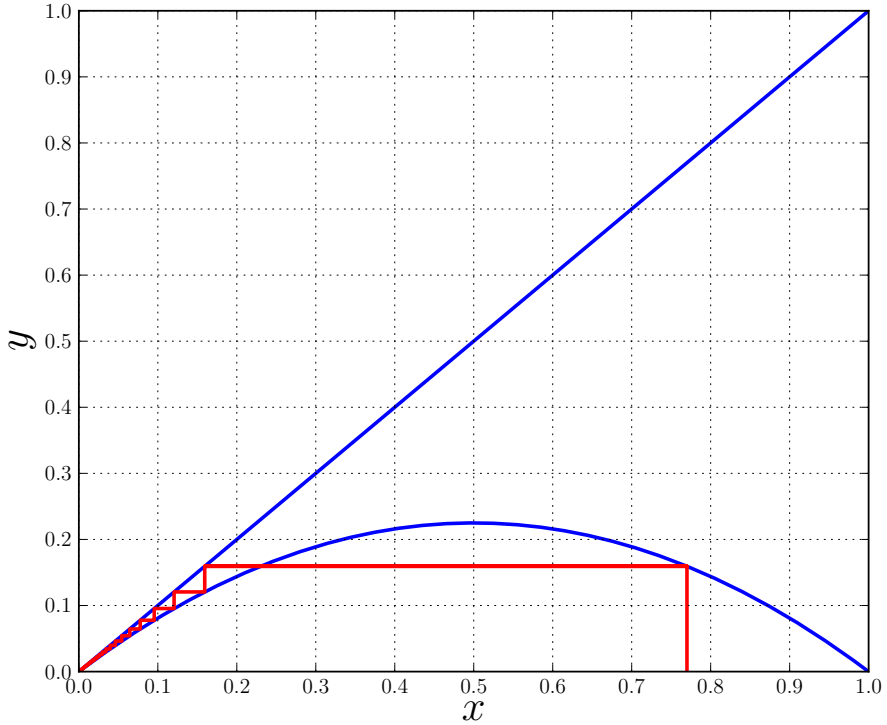
**EXERCISE** :

$$x_{k+1} \quad = \quad \lambda \, x_k \, (1 - x_k) \, , \qquad k = 1, 2, \; \cdots \; ,$$

- Divide the interval $[0, 1]$ into 200 *sub-intervals* .
- Compute $x_k$ for $k = 1, 2, \; \cdots \; 50$ .
- *Count* the $x_k$'s in each sub-interval.
- Determine the *percentage* of $x_k$'s in each sub-interval.
- Present the result in a *diagram* .
- Do this for different choices of $\lambda \quad (0 \leq \lambda \leq 4)$ .

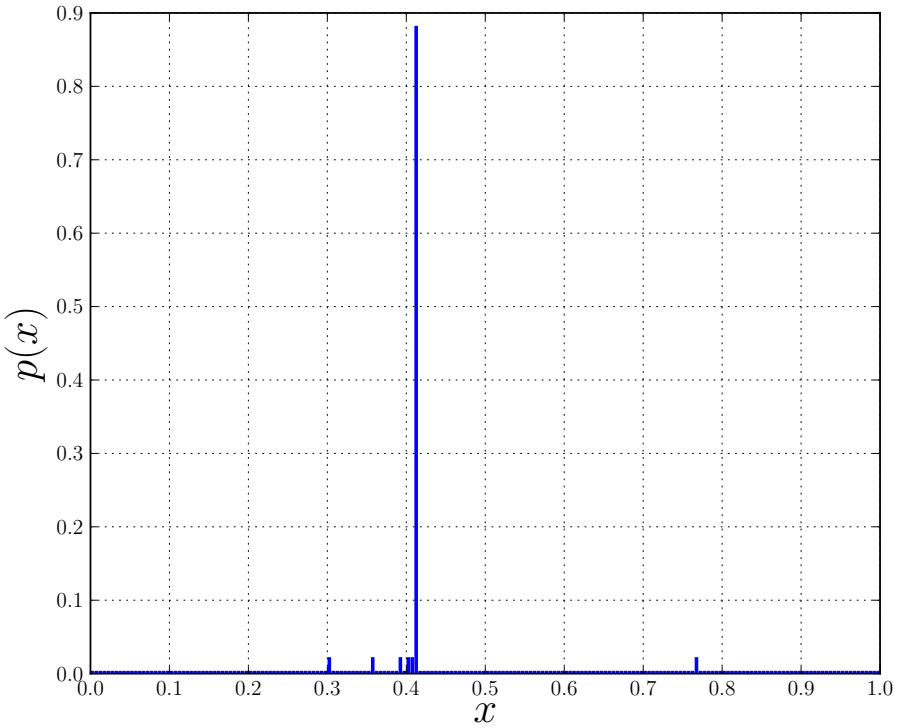



Percentage per interval

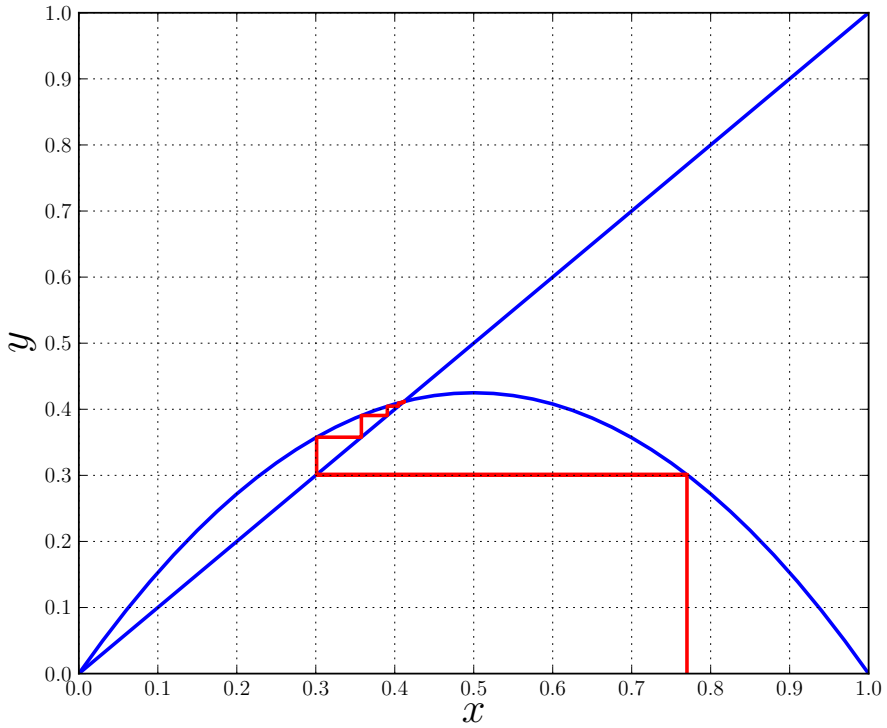



Graphical interpretation.

$\lambda = 0.9 \quad , \quad x_0 = 0.77 \quad , \quad 50$ iterations.

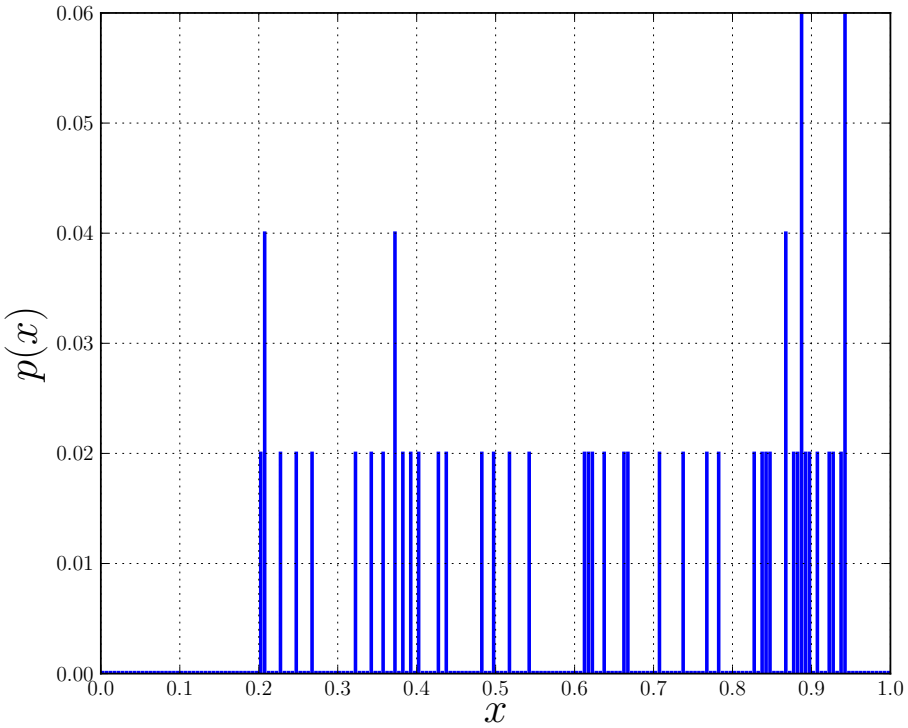



Percentage per interval

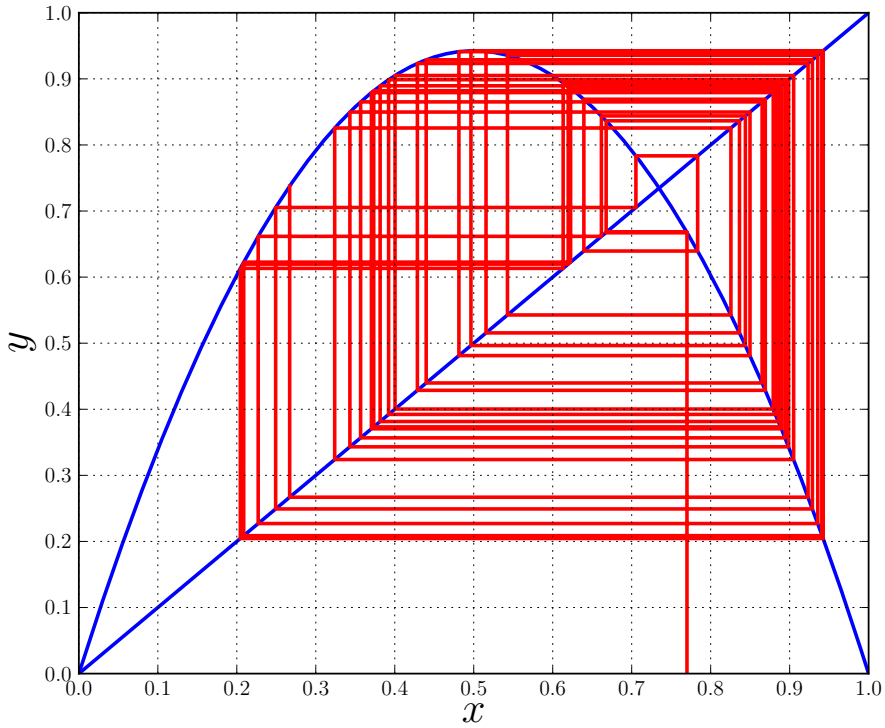



Graphical interpretation.

$\lambda = 1.7 \quad , \quad x_0 = 0.77 \quad , \quad 50$ iterations.




Percentage per interval




Graphical interpretation.

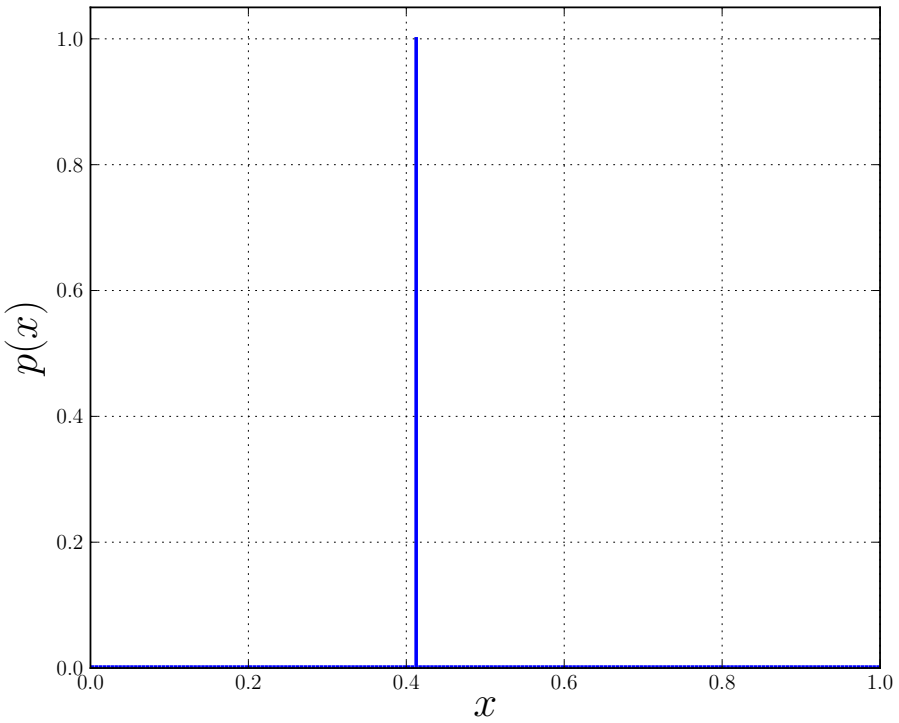
$\lambda = 3.77 \quad , \quad x_0 = 0.77 \quad , \quad 50$ iterations.

**EXERCISE** :

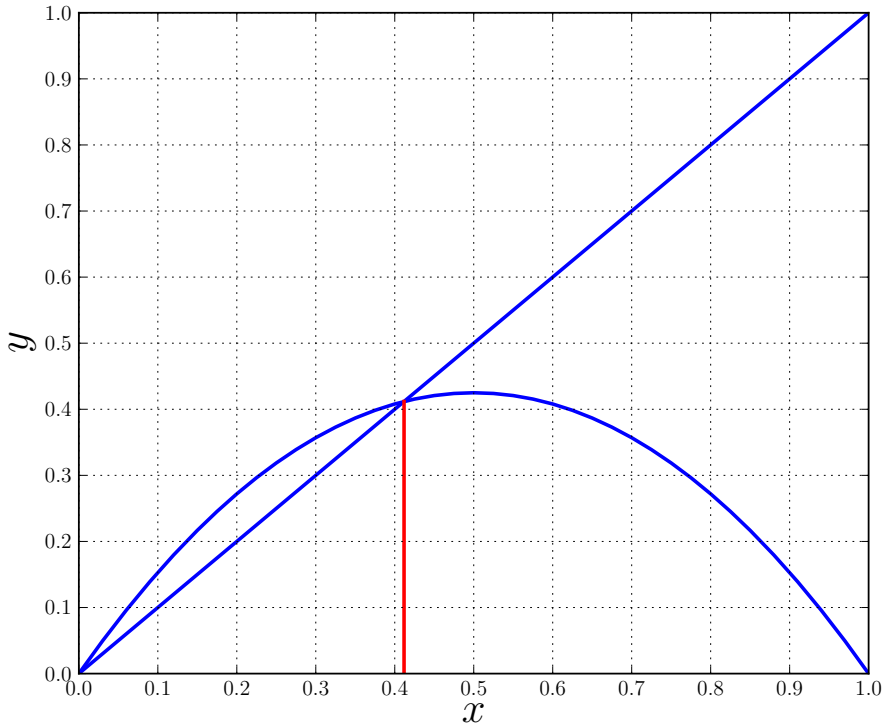
$$x_{k+1} \quad = \quad \lambda \, x_k \, (1 - x_k) \, , \qquad k = 1, 2, \; \cdots \; ,$$

Do the same as in the preceding example, but now

- Compute $x_k$ for $k = 1, 2, \; \cdots \; 1,000,000$ !
- Do not record the first 200 iterations.
- ( This to eliminate *transient effects* .)
- You will see that there is a *fixed point* (a *cycle of period 1* ) .



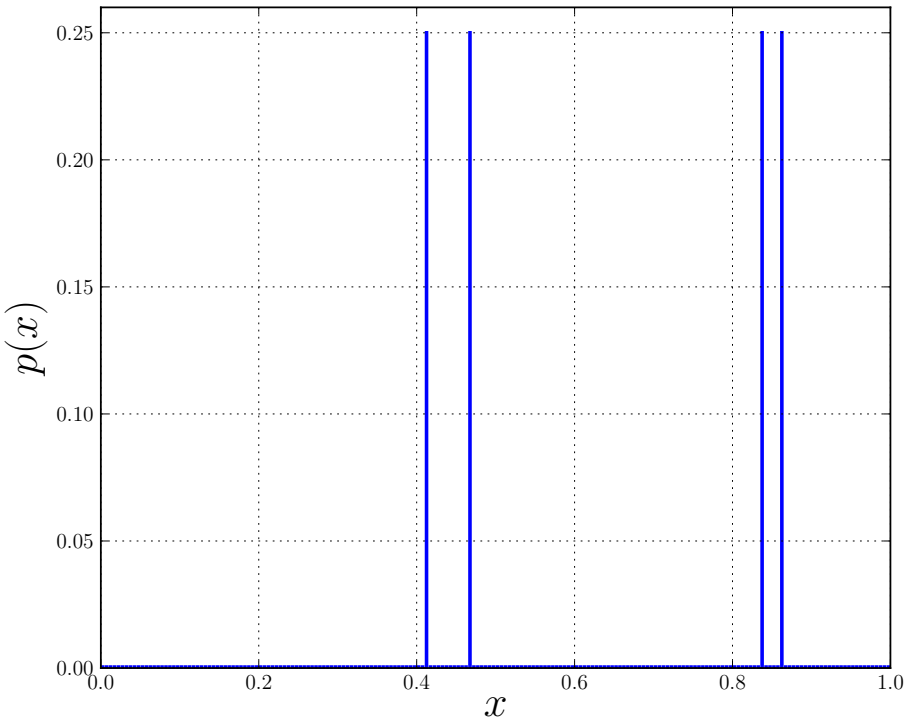

Percentage per interval



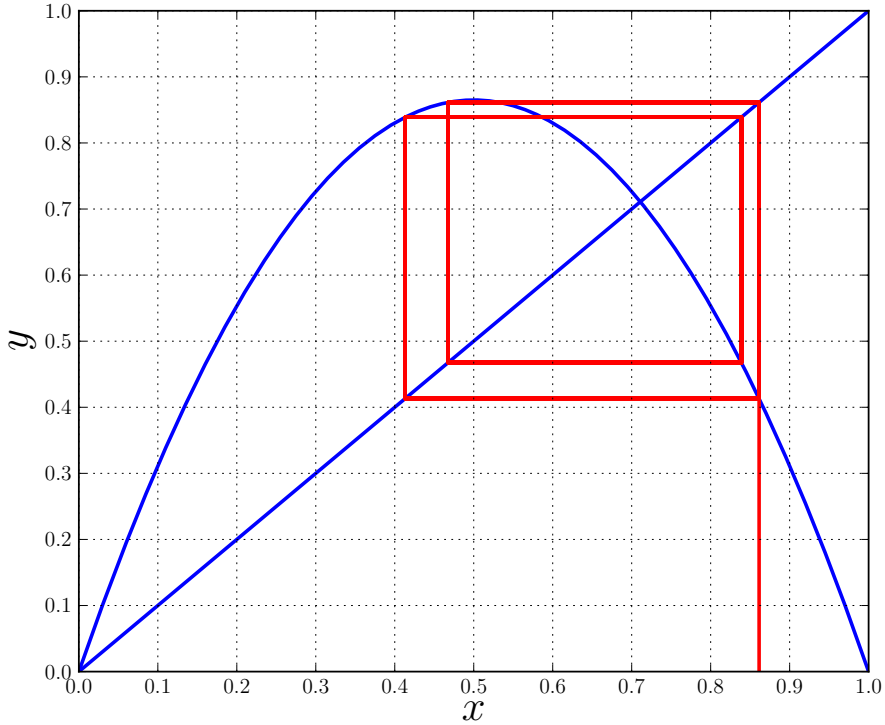

Graphical interpretation.

$\lambda = 1.7 \quad , \quad 1,000,000$ iterations.

There is a *fixed point* (a *cycle of period 1*) .



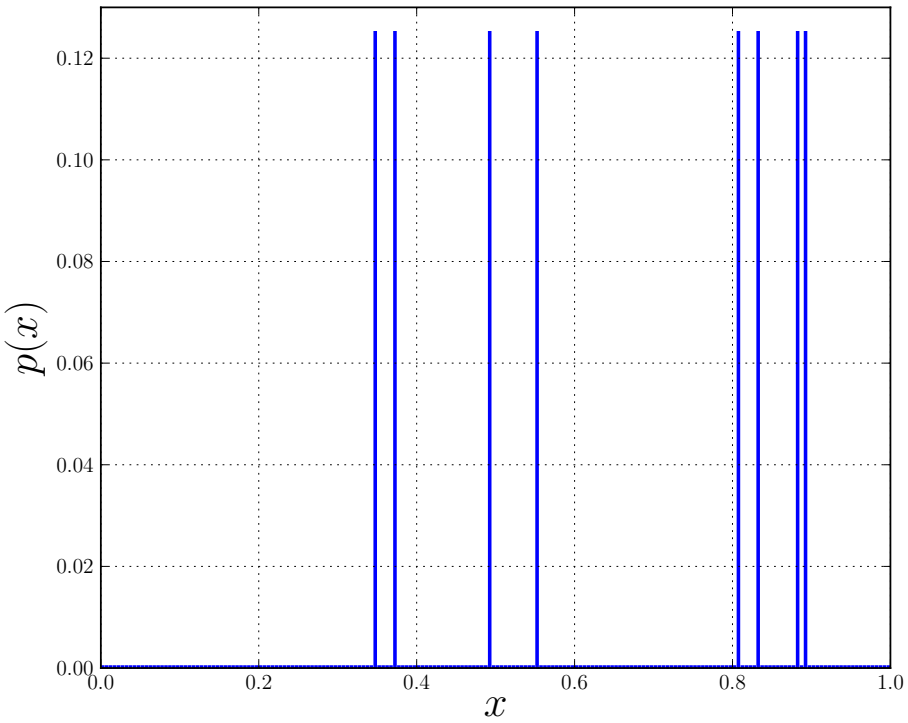

Percentage per interval



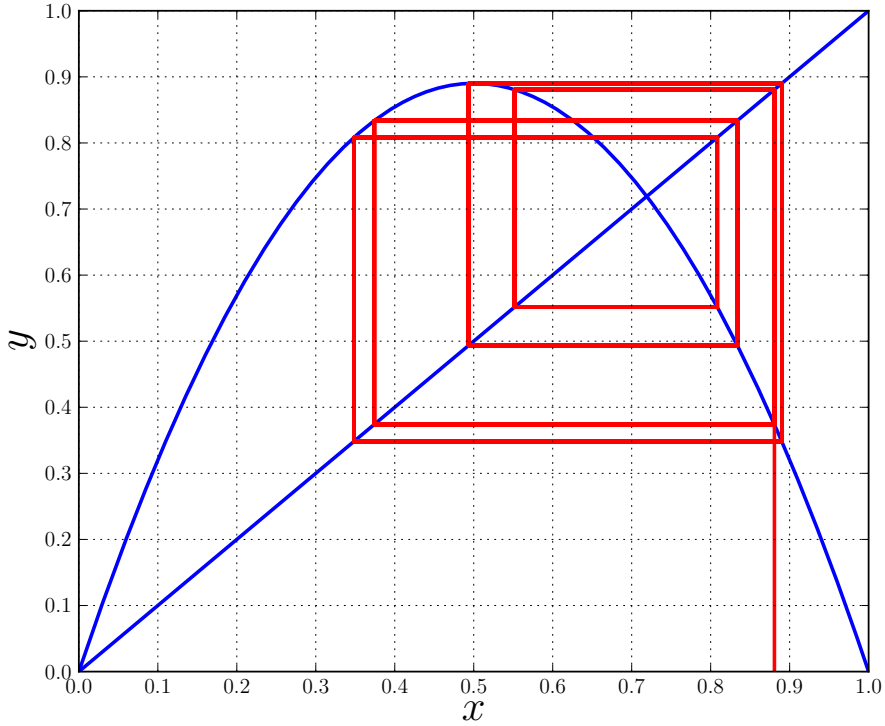

Graphical interpretation.

$\lambda = 3.46$ , $1,000,000$ iterations.

There is a *cycle of period 4* .



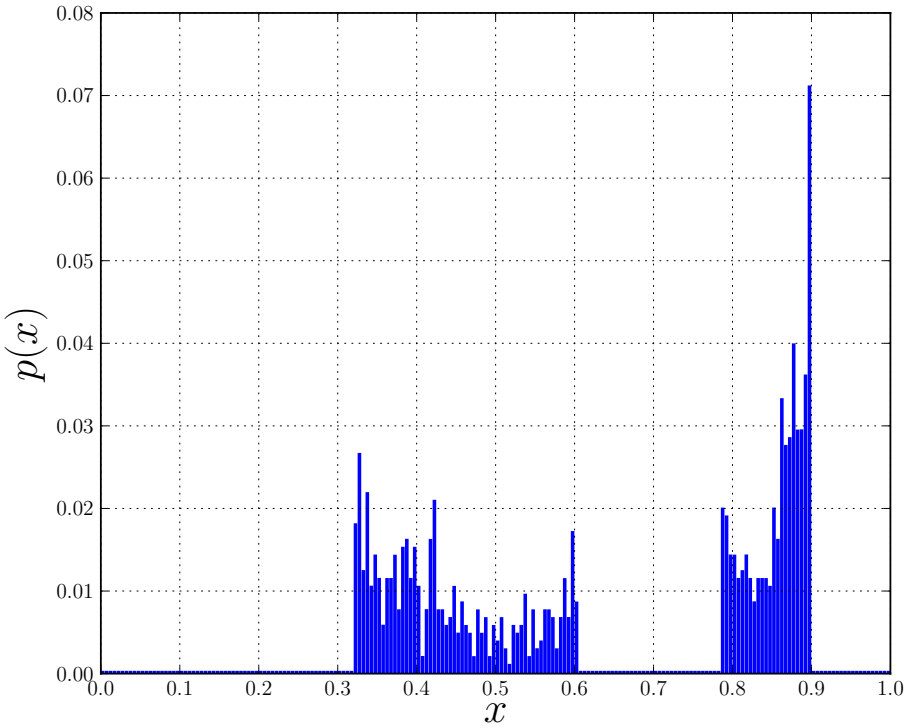

Percentage per interval



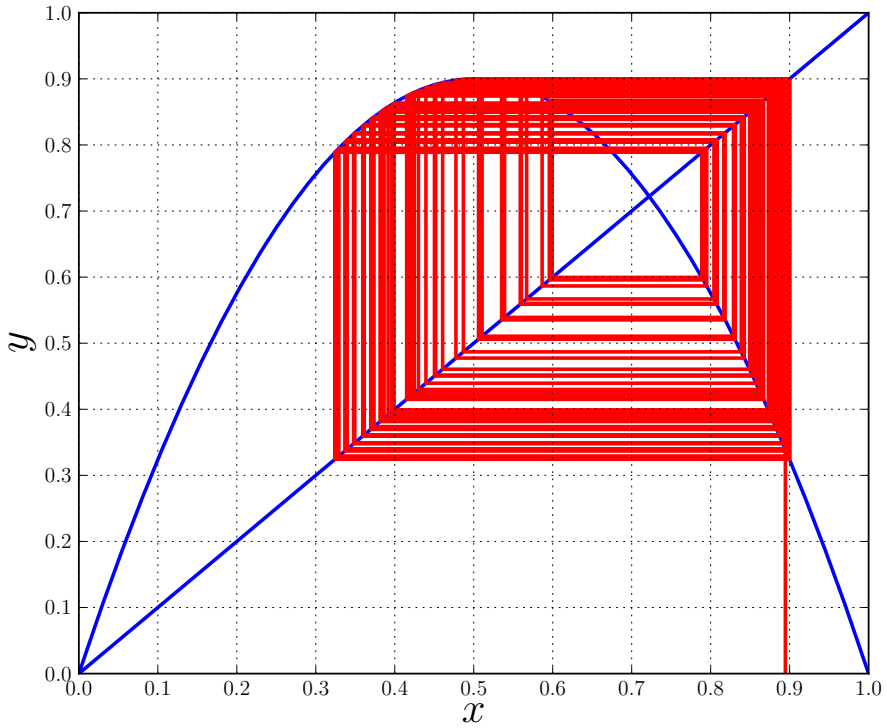

Graphical interpretation.

$\lambda = 3.561 \quad , \quad 1,000,000$ iterations.

There is a *cycle of period 8* .




Percentage per interval

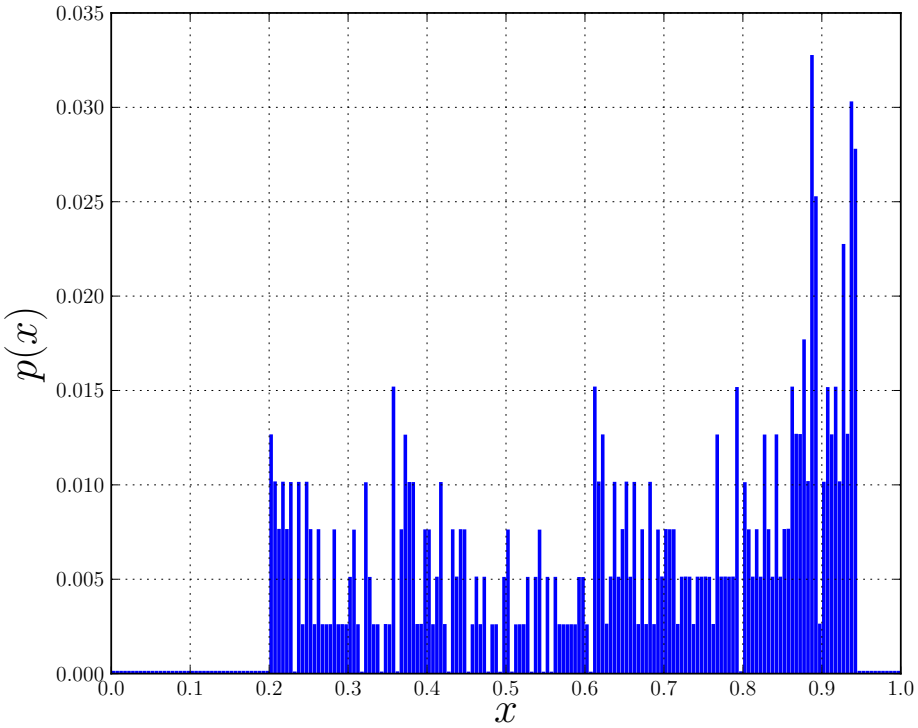


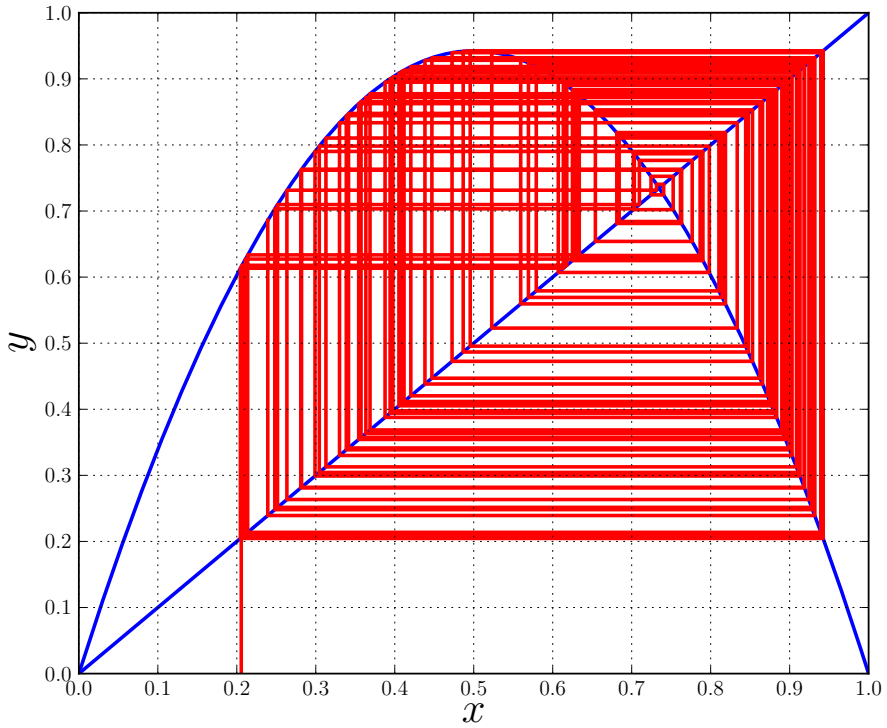

Graphical interpretation.

$\lambda = 3.6$ , $1,000,000$ iterations.

(The Figure on the right only shows 100 of these iterations.)

There is apparent *chaotic* behavior .



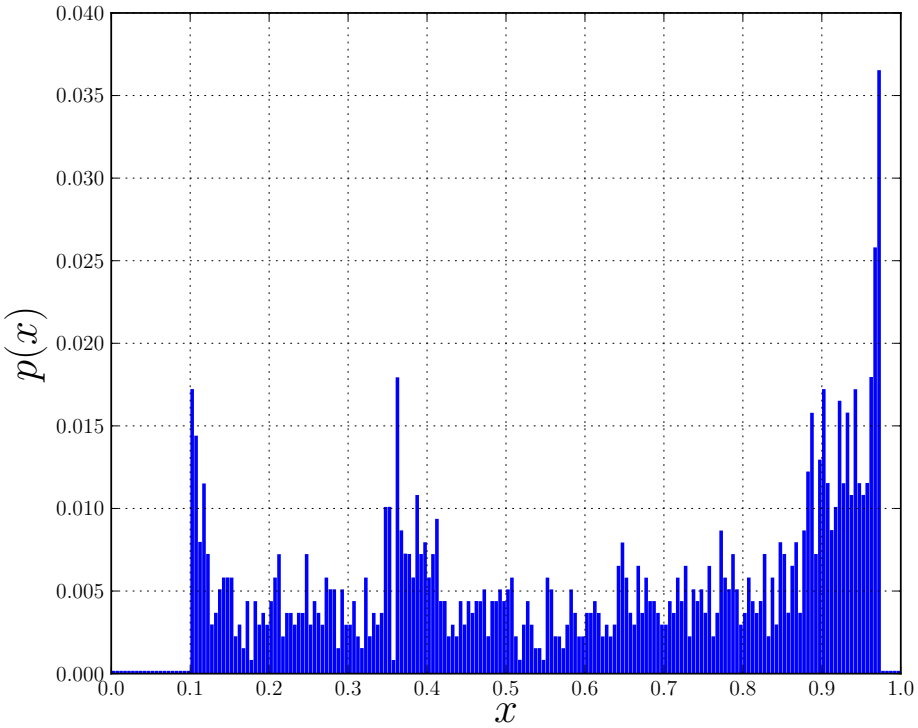

Percentage per interval

Graphical interpretation.

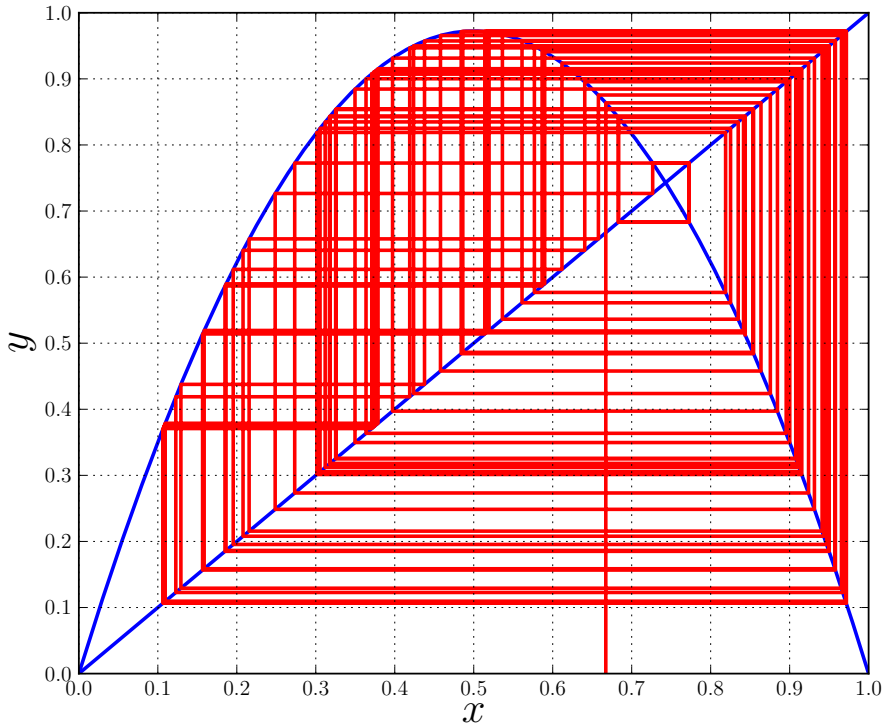
$\lambda = 3.77$ , $1,000,000$ iterations.

(The Figure on the right only shows 100 of these iterations.)

There is apparent *chaotic* behavior .




Percentage per interval

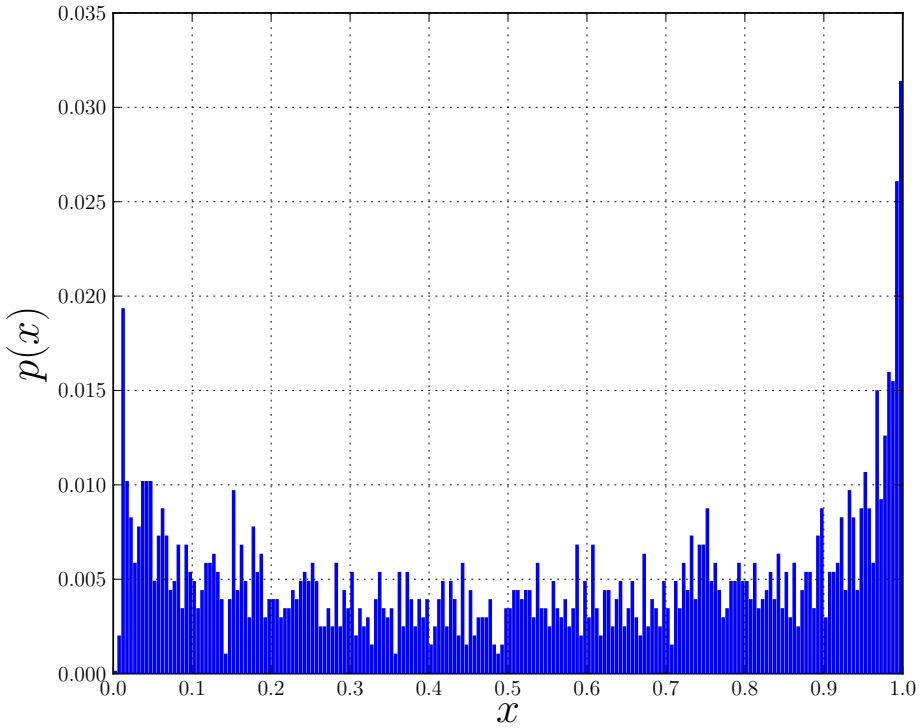



Graphical interpretation.

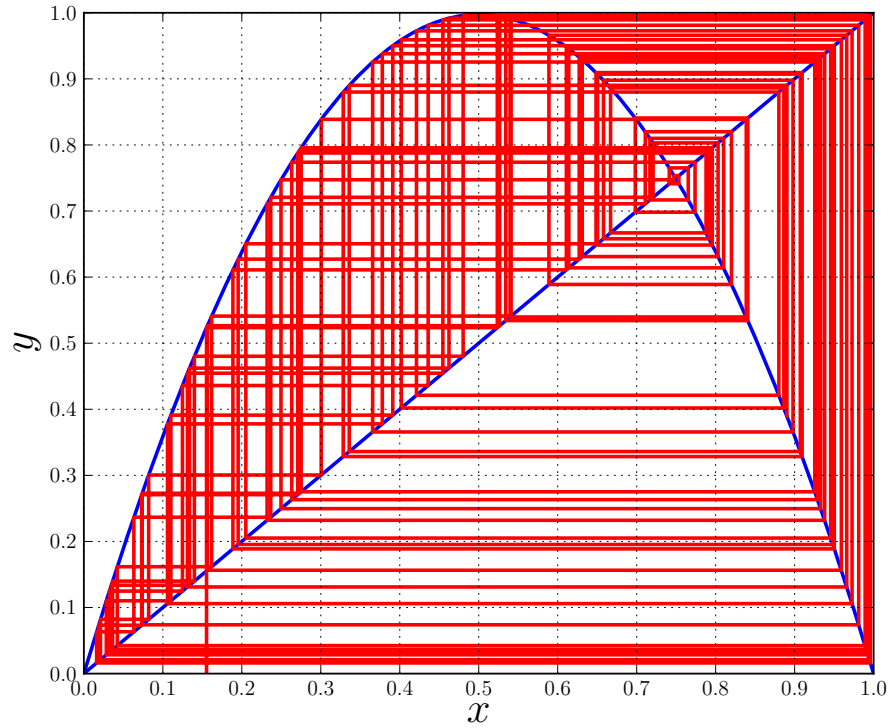
$\lambda = 3.89$ , $1,000,000$ iterations.

(The Figure on the right only shows 100 of these iterations.)

There is apparent *chaotic* behavior .




Percentage per interval




Graphical interpretation.

$\lambda = 3.99$ , $1,000,000$ iterations.

(The Figure on the right only shows 100 of these iterations.)

There is apparent *chaotic* behavior .

## CONCLUSIONS :

- The behavior of the logistic equation depends on $\lambda$ .
- For certain values of $\lambda$ we see *fixed points* .
- For other values of $\lambda$ there are *cycles* .
- For yet other values of $\lambda$ there is seemingly *random behavior* .
- Many other deterministic equations have "*complex behavior*" !
- Nature is complex !

## Generating Random Numbers

The logistic equation is a *recurrence relation* of the form

$$x_{k+1} \quad = \quad f(x_k) \ , \quad k = 1, 2, 3, \ \cdots \ .$$

Most random number generators also have this form.

For the logistic equation we did *not* see sequences $\{x_k\}_{k=1}^{N}$ having

- a uniform distribution,
- a normal distribution,
- any other known distribution.

**QUESTION** :

- How to generate *uniform* (and other) random numbers ?
- ( These are useful in *computer simulations* .)

## Generating Uniformly Distributed Random Numbers

Uniformly distributed random numbers can also be generated by a *recurrence relation* of the form

$$x_{k+1} \quad = \quad f(x_k) \ , \quad k = 1, 2, 3, \ \cdots \ .$$

Unlike the logistic equation, the $x_k$ are most often *integers* .

The recurrence relation typically has the form

$$x_{k+1} \quad = \quad (n \ x_k) \bmod p \ .$$

where $p$ is a prime number, and $n$ an integer such that

$$p \quad \not| \ n \ .$$

The following fact is useful :

**THEOREM** :

Let $p$ be a prime number, and $n$ an integer such that

$$p \not| n \ .$$

Then the function

$$f \quad : \quad \{0, 1, 2, \cdots, p-1\} \quad \rightarrow \quad \{0, 1, 2, \cdots, p-1\} \ ,$$

given by

$$f(x) \ = \ (n \ x) \bmod p \ ,$$

is *one-to-one* (and hence *onto*, a *bijection*, and *invertible*).

$$\boxed{p \text{ prime}, \; p \nmid n \quad \Rightarrow \quad (n\ x) \bmod p \text{ is } 1-1}$$

**EXAMPLE** :

$$p = 7 \quad \text{and} \quad n = 12 \, .$$

| x | 12x | 12x mod 7 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 12 | 5 |
| 2 | 24 | 3 |
| 3 | 36 | 1 |
| 4 | 48 | 6 |
| 5 | 60 | 4 |
| 6 | 72 | 2 |

*Invertible* !

**NOTE** : The numbers in the right hand column look rather *random* !

$$\boxed{p \text{ prime}\,,\ p \not| n \quad \Rightarrow \quad (n\ x) \bmod p \quad \text{is} \ \ 1-1}$$

**EXAMPLE** :

$$p \ = \ 6 \quad \text{and} \quad n \ = \ 2\,.$$

| x | 2x | 2x mod 6 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 2 | 2 |
| 2 | 4 | 4 |
| 3 | 6 | 0 |
| 4 | 8 | 2 |
| 5 | 10 | 4 |

*Not Invertible* .

$$\boxed{p \text{ prime}, \; p \nmid n \quad \Rightarrow \quad (n\ x) \bmod p \text{ is } 1-1}$$

**EXAMPLE** :

$$p = 6 \quad \text{and} \quad n = 13 \, .$$

| x | 13x | 13x mod 6 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 13 | 1 |
| 2 | 26 | 2 |
| 3 | 39 | 3 |
| 4 | 52 | 4 |
| 5 | 65 | 5 |

*Invertible* . ( So ? )

**NOTE** : The numbers in the right hand column don't look *random* !

$$\boxed{p \text{ prime }, \ p \not| \, n \quad \Rightarrow \quad (n\ x) \bmod p \quad \text{is} \ \ 1-1}$$

**PROOF** : *By contradiction* : Suppose the function is *not* $1-1$ .

Then there are *distinct* integers $x_1$ and $x_2$ , such that

$$(nx_1) \bmod p \ = \ k \qquad \text{and} \qquad (nx_2) \bmod p \ = \ k \ ,$$

where

$$x_1 \ , \ x_2 \ , \ k \ \in \ \{0, 1, 2, \cdots, p-1\} \ .$$

It follows that

$$p \ | \ n(x_1 - x_2) \ , \qquad (\text{ Why ? })$$

Since $p$ is prime and $p \not| \, n$ it follows that

$$p \mid (x_1 - x_2) \ .$$

Thus $x_1 - x_2 \ = \ 0$ ( Why ? ) , *i.e.* , $x_1 \ = \ x_2$ . Contradiction!

For given $x_0$ , an iteration of the form

$$x_k \;=\; (n\; x_{k-1}) \bmod p \;, \qquad k \;=\; 1, 2, \;\cdots\; , p-1 \;.$$

can be used to generate *random numbers* .

- Here $p$ is a *large prime number* .
- The value of $n$ is also *large* .
- The integer $n$ must *not* be divisible by $p$ .
- Do not start with $x_0 = 0$ (because it is a *fixed point* !) .
- Be aware of *cycles* (of period less than $p-1$) !

**REMARK** : Actually, *more often used* is an iteration of the form

$$x_k \;=\; (n\; x_{k-1} + m) \bmod p \;, \qquad k \;=\; 1, 2, \;\cdots\; , p-1 \;.$$

**EXAMPLE** : As a simple example, take again $p = 7$ and $n = 12$ :

| x | 12x | 12x mod 7 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 12 | 5 |
| 2 | 24 | 3 |
| 3 | 36 | 1 |
| 4 | 48 | 6 |
| 5 | 60 | 4 |
| 6 | 72 | 2 |

With $x_1 = 1$ the recurrence relation

$$x_{k+1} \;=\; f(x_k)\;, \qquad \text{where} \quad f(x) \;=\; 12x \bmod 7\;,$$

generates the sequence

$$1 \;\to\; 5 \;\to\; 4 \;\to\; 6 \;\to\; 2 \;\to\; 3 \;\to\; 1 \;\to\; 5 \;\to\; \cdots$$

which is *a cycle of maximal period* $p-1$ (here 6) .

| $f(x)$ | sequence | cycle period |
|---|---|---|
| $5x \bmod 7$ | $1 \to 5 \to 4 \to 6 \to 2 \to 3 \to 1$ | **6** |
| $6x \bmod 7$ | $1 \to 6 \to 1 \to 6 \to 1 \to 6 \to 1$ | **2** |
| $8x \bmod 7$ | $1 \to 1 \to 1 \to 1 \to 1 \to 1 \to 1$ | **1** |
| $9x \bmod 7$ | $1 \to 2 \to 4 \to 1 \to 2 \to 4 \to 1$ | **3** |
| $10x \bmod 7$ | $1 \to 3 \to 2 \to 6 \to 4 \to 5 \to 1$ | **6** |
| $11x \bmod 7$ | $1 \to 4 \to 2 \to 1 \to 4 \to 2 \to 1$ | **3** |
| $12x \bmod 7$ | $1 \to 5 \to 4 \to 6 \to 2 \to 3 \to 1$ | **6** |

**EXAMPLE** : With $x_0 = 2$ , compute

$$x_k \;=\; (137951\; x_{k-1}) \bmod 101 \;, \qquad k \;=\; 1, 2, \;\cdots\;, 100 \;.$$

Result :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 71 | 46 | 17 | 48 | 88 | 94 | 4 | 41 | 92 | 34 |
| 96 | 75 | 87 | 8 | 82 | 83 | 68 | 91 | 49 | 73 |
| 16 | 63 | 65 | 35 | 81 | 98 | 45 | 32 | 25 | 29 |
| 70 | 61 | 95 | 90 | 64 | 50 | 58 | 39 | 21 | 89 |
| 79 | 27 | 100 | 15 | 78 | 42 | 77 | 57 | 54 | 99 |
| 30 | 55 | 84 | 53 | 13 | 7 | 97 | 60 | 9 | 67 |
| 5 | 26 | 14 | 93 | 19 | 18 | 33 | 10 | 52 | 28 |
| 85 | 38 | 36 | 66 | 20 | 3 | 56 | 69 | 76 | 72 |
| 31 | 40 | 6 | 11 | 37 | 51 | 43 | 62 | 80 | 12 |
| 22 | 74 | 1 | 86 | 23 | 59 | 24 | 44 | 47 | 2 |

**QUESTION** : Are there *repetitions* (*i.e.*, *cycles*) ?

**EXAMPLE** : As in the preceding example, use $x_0 = 2$ , and compute

$$x_k = (137951\ x_{k-1}) \bmod 101 , \qquad k = 1, 2, \cdots , 100 ,$$

and set

$$\hat{x}_k = \frac{x_k}{100} .$$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.710 | 0.460 | 0.170 | 0.480 | 0.880 | 0.940 | 0.040 | 0.410 | 0.920 | 0.340 |
| 0.960 | 0.750 | 0.870 | 0.080 | 0.820 | 0.830 | 0.680 | 0.910 | 0.490 | 0.730 |
| 0.160 | 0.630 | 0.650 | 0.350 | 0.810 | 0.980 | 0.450 | 0.320 | 0.250 | 0.290 |
| 0.700 | 0.610 | 0.950 | 0.900 | 0.640 | 0.500 | 0.580 | 0.390 | 0.210 | 0.890 |
| 0.790 | 0.270 | 1.000 | 0.150 | 0.780 | 0.420 | 0.770 | 0.570 | 0.540 | 0.990 |
| 0.300 | 0.550 | 0.840 | 0.530 | 0.130 | 0.070 | 0.970 | 0.600 | 0.090 | 0.670 |
| 0.050 | 0.260 | 0.140 | 0.930 | 0.190 | 0.180 | 0.330 | 0.100 | 0.520 | 0.280 |
| 0.850 | 0.380 | 0.360 | 0.660 | 0.200 | 0.030 | 0.560 | 0.690 | 0.760 | 0.720 |
| 0.310 | 0.400 | 0.060 | 0.110 | 0.370 | 0.510 | 0.430 | 0.620 | 0.800 | 0.120 |
| 0.220 | 0.740 | 0.010 | 0.860 | 0.230 | 0.590 | 0.240 | 0.440 | 0.470 | 0.020 |

**QUESTION** : Do these numbers look *uniformly distributed* ?

**EXAMPLE** : With $x_0 = \mathbf{2}$ , compute

$$x_k \;=\; (137953\; x_{k-1}) \bmod 101 \;, \qquad k \;=\; 1, 2, \;\cdots\;, 100 \;.$$

Result :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 75 | 35 | 50 | 57 | 67 | 38 | 11 | 59 | 41 | 73 |
| 61 | 15 | 7 | 10 | 72 | 74 | 48 | 83 | 32 | 89 |
| 55 | 93 | 3 | 62 | 2 | 75 | 35 | 50 | 57 | 67 |
| 38 | 11 | 59 | 41 | 73 | 61 | 15 | 7 | 10 | 72 |
| 74 | 48 | 83 | 32 | 89 | 55 | 93 | 3 | 62 | 2 |
| 75 | 35 | 50 | 57 | 67 | 38 | 11 | 59 | 41 | 73 |
| 61 | 15 | 7 | 10 | 72 | 74 | 48 | 83 | 32 | 89 |
| 55 | 93 | 3 | 62 | 2 | 75 | 35 | 50 | 57 | 67 |
| 38 | 11 | 59 | 41 | 73 | 61 | 15 | 7 | 10 | 72 |
| 74 | 48 | 83 | 32 | 89 | 55 | 93 | 3 | 62 | 2 |

**QUESTION** : Are there *cycles* ?

**EXAMPLE** : With $x_0 = 4$ , compute

$$x_k \;=\; (137953\, x_{k-1}) \bmod 101 \;, \qquad k \;=\; 1, 2, \;\cdots\;, 100 \;.$$

Result :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 49 | 70 | 100 | 13 | 33 | 76 | 22 | 17 | 82 | 45 |
| 21 | 30 | 14 | 20 | 43 | 47 | 96 | 65 | 64 | 77 |
| 9 | 85 | 6 | 23 | 4 | 49 | 70 | 100 | 13 | 33 |
| 76 | 22 | 17 | 82 | 45 | 21 | 30 | 14 | 20 | 43 |
| 47 | 96 | 65 | 64 | 77 | 9 | 85 | 6 | 23 | 4 |
| 49 | 70 | 100 | 13 | 33 | 76 | 22 | 17 | 82 | 45 |
| 21 | 30 | 14 | 20 | 43 | 47 | 96 | 65 | 64 | 77 |
| 9 | 85 | 6 | 23 | 4 | 49 | 70 | 100 | 13 | 33 |
| 76 | 22 | 17 | 82 | 45 | 21 | 30 | 14 | 20 | 43 |
| 47 | 96 | 65 | 64 | 77 | 9 | 85 | 6 | 23 | 4 |

**QUESTIONS** :

- Are there *cycles* ?
- Is this the *same cycle* that we already found ?

## Generating Random Numbers using the Inverse Method

- There are algorithms that generate *uniform* random numbers.
- These can be used to generate *other* random numbers.
- A simple method to do this is the *Inverse Transform Method* .

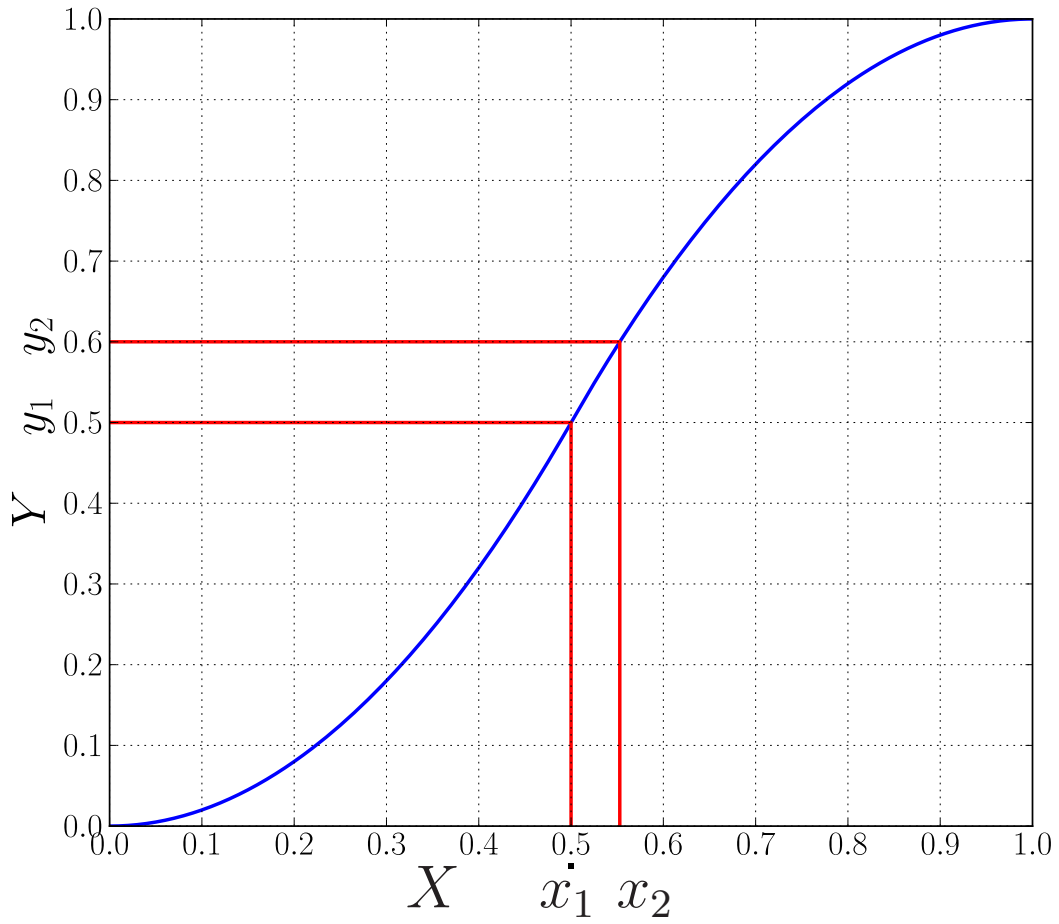
**RECALL** :

Let $f(x)$ be a *density function* on an interval $[a, b]$ .

The *distribution function* is

$$F(x) \equiv \int_a^x f(x)\, dx \ .$$

- Since $f(x) \geq 0$ we know that $F(x)$ is *increasing* .
- If $F(x)$ is *strictly increasing* then $F(x)$ is *invertible* .




We want random numbers $X$ with distribution $F(x)$ (blue) .

Let the random variable $Y$ be uniform on the interval $[0, 1]$ .

$$\text{Let} \quad X \;=\; F^{-1}(Y) \;.$$

$$\text{Then} \quad P(x_1 \le X \le x_2) \;=\; y_2 - y_1 \;=\; F(x_2) - F(x_1) \;.$$

Thus $F(x)$ is indeed the distribution function of $X$ !

- If $Y$ is uniformly distributed on $[0,1]$ then

$$P(\ y_1 \le Y \le y_2\ ) \quad = \quad y_2 - y_1\ .$$

- Let

$$X \quad = \quad F^{-1}(Y)\ ,$$

with

$$x_1 \ = \ F^{-1}(y_1) \qquad \text{and} \qquad x_2 \ = \ F^{-1}(y_2)\ .$$

- Then

$$P(x_1 \le X \le x_2) \ = \ y_2 - y_1 \ = \ F(x_2) - F(x_1)\ . \qquad (\text{ Why ? })$$

- Thus $F(X)$ is also the *distribution function* of $X \equiv F^{-1}(Y)$ !

**NOTE** : In the illustration $X$ is on $[0,1]$ , but this is not necessary.

**EXAMPLE** :

Recall that the *exponential density function*

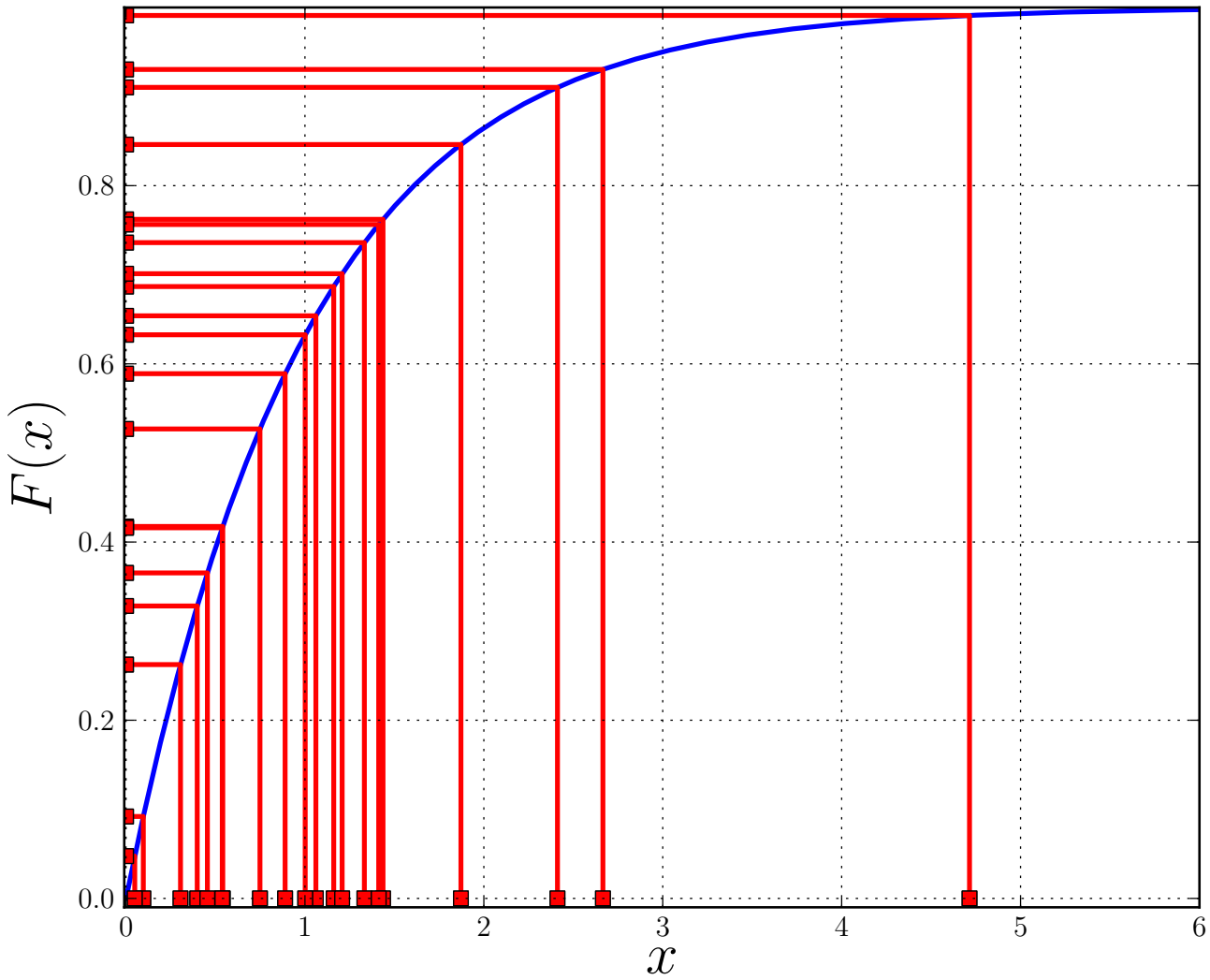
$$f(x) \;=\; \begin{cases} \lambda e^{-\lambda x} \;, & x > 0 \\ \\ 0 \;, & x \leq 0 \end{cases}$$

has *distribution function*

$$F(x) \;=\; \begin{cases} 1 - e^{-\lambda x} \;, & x > 0 \\ \\ 0 \;, & x \leq 0 \end{cases}$$

The *inverse distribution function* is

$$F^{-1}(y) \;=\; -\frac{1}{\lambda} \log(1-y) \;, \qquad 0 \leq y < 1 \;. \qquad (\text{ Check ! })$$



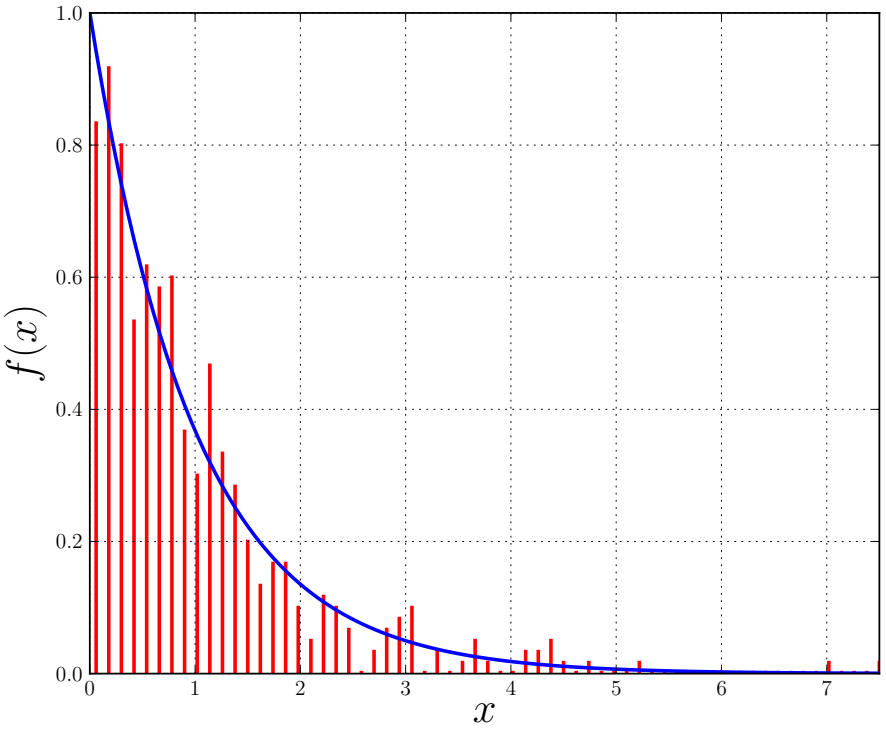

The inverse method for generating 20 random numbers.

(Shown for the *exponential distribution function* with $\lambda = 1$ .)

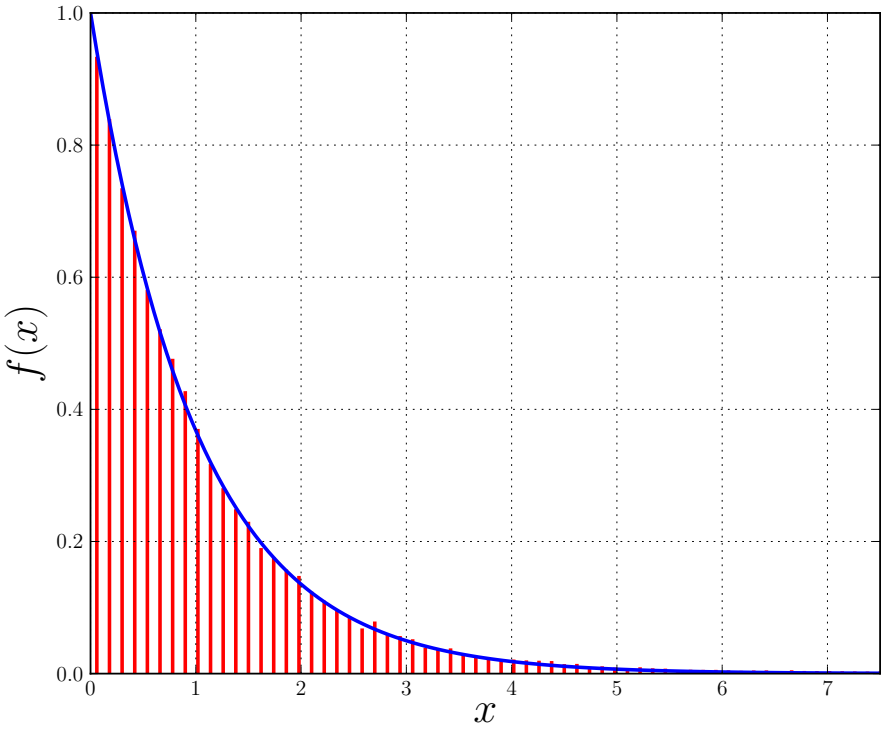
**EXAMPLE** : ( continued $\cdots$ )

(Using the *inverse method* to get *exponential* random numbers.)

- Divide $[0, 12]$ into 100 subintervals of equal size $\Delta x$ .
- Let $I_k$ denote the $k$th interval, with midpoint $x_k$ .
- Use the *inverse method* to get $N$ *exponential* random numbers.
- Let $m_k$ be the *frequency count* (# of random values in $I_k$) .
- Let $$\hat{f}(x_k) \quad = \quad \frac{m_k}{N \ \Delta x} \ .$$
- Then $$\int_0^\infty \hat{f}(x) \ dx \quad \cong \quad \sum_{k=1}^{100} \hat{f}(x_k) \ \Delta x \quad = \quad 1 \ .$$
- Thus $\hat{f}(x)$ should approximate the actual *density function* $f(x)$ .




$N = 500$ .




$N = 50,000$ .

Simulating the exponential random variable with $\lambda = 1$ .

(The actual density function is shown in *blue*.)

**EXERCISE** : Consider the *Tent density function*

$$f(x) \;=\; \begin{cases} x+1\,, & -1 < x \leq 0 \\ 1-x\,, & 0 < x \leq 1 \\ 0\,, & \text{otherwise} \end{cases}$$

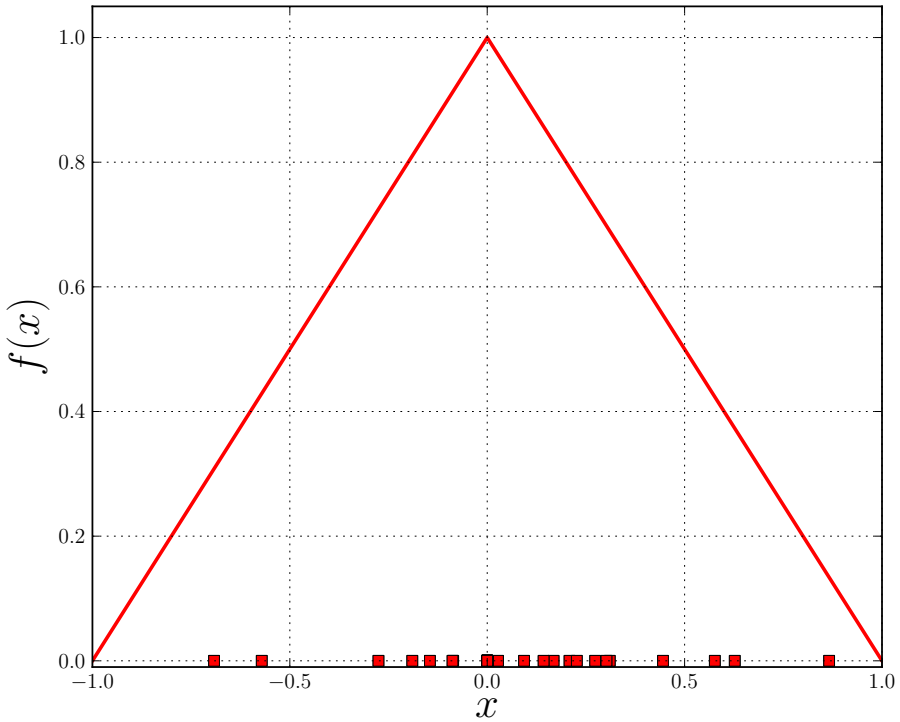
- Verify that that the distribution function is given by

$$F(x) \;=\; \begin{cases} \frac{1}{2}x^2 + x + \frac{1}{2}\,, & -1 \leq x \leq 0 \\ \\ -\frac{1}{2}x^2 + x + \frac{1}{2}\,, & 0 < x \leq 1 \end{cases}$$

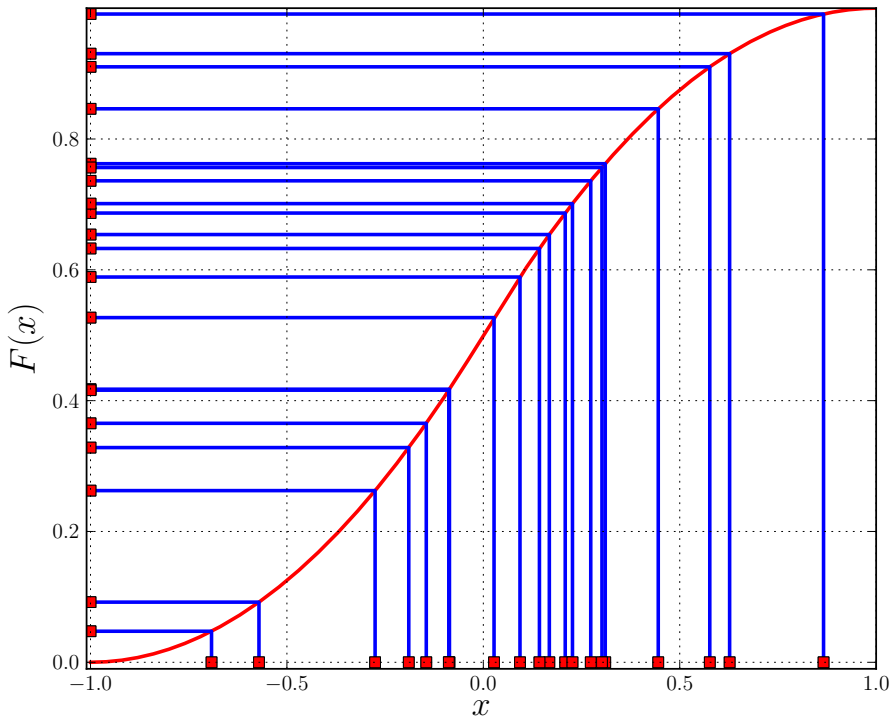
- Verify that that the *inverse* distribution function is

$$F^{-1}(y) \;=\; \begin{cases} -1 + \sqrt{2y}\,, & 0 \leq y \leq \frac{1}{2} \\ \\ 1 - \sqrt{2-2y}\,, & \frac{1}{2} < y \leq 1 \end{cases}$$

- Use the inverse method to generate " *Tent random numbers* ".

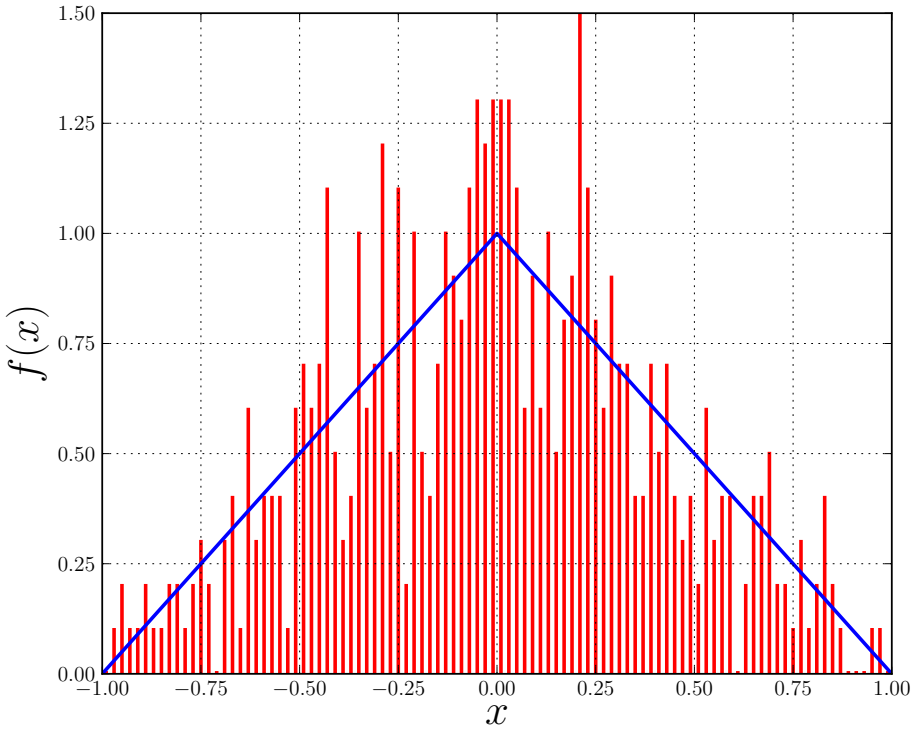


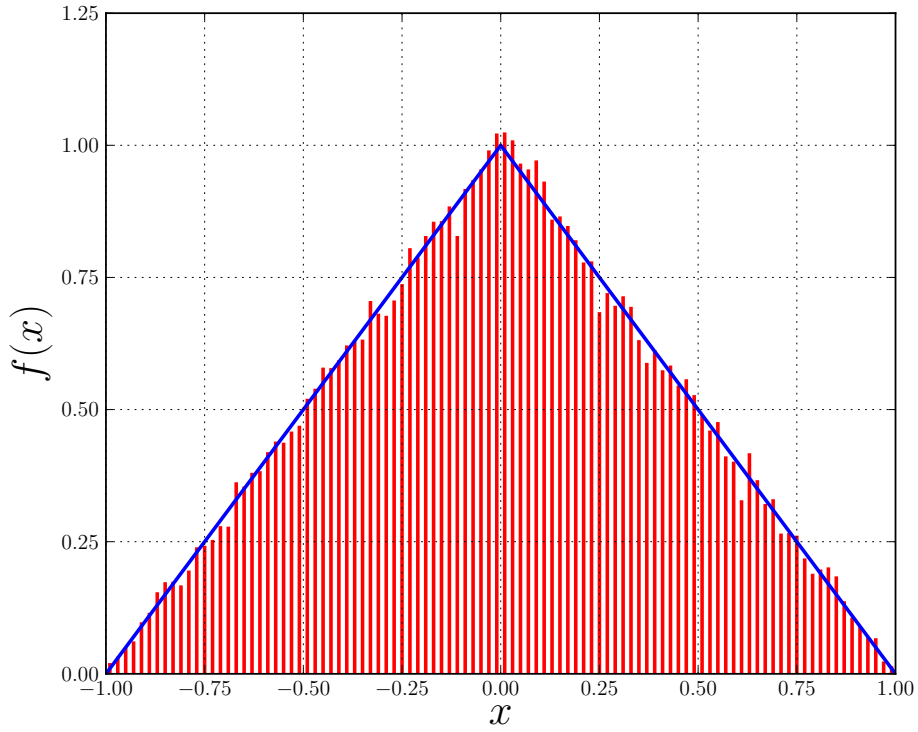

The "*Tent*" Density Function.




Distribution function.

Using the inverse method to generate 20 "*Tent random values*".




$N = 500$ .

$N = 50,000$ .

Simulating the "*Tent random variable*" .

(The actual density function is shown in *blue*.)

# SUMMARY TABLES

# and

# FORMULAS

| Discrete | Continuous |
|---|---|
| $p(x_i) = P(X = x_i)$ | $f(x)\delta x \;\cong\; P(x - \frac{\delta}{2} < X < x + \frac{\delta}{2})$ |
| $\sum_i p(x_i) = 1$ | $\int_{-\infty}^{\infty} f(x)\ dx = 1$ |
| $F(x_k) = \sum_{i \le k} p(x_i)$ | $F(x) = \int_{-\infty}^{x} f(x)\ dx$ |
| $p(x_k) = F(x_k) - F(x_{k-1})$ | $f(x) = F'(x)$ |
| $E[X] = \sum_i x_i p(x_i)$ | $E[X] = \int_{-\infty}^{\infty} x\ f(x)\ dx$ |
| $E[g(X)] = \sum_i g(x_i) p(x_i)$ | $E[g(X)] = \int_{-\infty}^{\infty} g(x)\ f(x)\ dx$ |
| $E[XY] = \sum_{i,j} x_i y_j p(x_i, y_j)$ | $E[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y)\ dy\ dx$ |

| Name | General Formula |
|---|---|
| Mean | $\mu = E[X]$ |
| Variance | $Var(X) \;=\; E[(X-\mu)^2] \;=\; E[X^2] - \mu^2$ |
| Covariance | $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X \mu_Y$ |
| Markov | $P(X \geq c) \;\leq\; E[X]/c$ |
| Chebyshev | $P(\mid X - \mu \mid \geq k\sigma) \;\leq\; 1/k^2$ |
| Moments | $\psi(t) = E[e^{tX}] \quad , \quad \psi'(0) = E[X] \quad , \quad \psi''(0) = E[X^2]$ |

| **Name** | **Probability mass function** | **Domain** |
|---|---|---|
| Bernoulli | $P(X = 1) = p \ , \ P(X = 0) = 1 - p$ | $0 \ , \ 1$ |
| Binomial | $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ | $0 \leq k \leq n$ |
| Poisson | $P(X = k) \ = \ e^{-\lambda} \ \frac{\lambda^k}{k!}$ | $k = 0, 1, 2, \cdots$ |

| Name | Mean | Standard deviation |
|---|---|---|
| Bernoulli | $p$ | $\sqrt{p(1-p)}$ |
| Binomial | $np$ | $\sqrt{np(1-p)}$ |
| Poisson | $\lambda$ | $\sqrt{\lambda}$ |

| Name | Density function | Distribution | Domain |
|---|---|---|---|
| Uniform | $\frac{1}{b-a}$ | $\frac{x-a}{b-a}$ | $x \in (a, b]$ |
| Exponential | $\lambda e^{-\lambda x}$ | $1 - e^{-\lambda x}$ | $x \in (0, \infty)$ |
| Std. Normal | $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ | | $x \in (-\infty, \infty)$ |
| Normal | $\frac{1}{\sqrt{2\pi}\ \sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$ | | $x \in (-\infty, \infty)$ |

| Name | Mean | Standard Deviation |
|---|---|---|
| Uniform | $\frac{a+b}{2}$ | $\frac{b-a}{2\sqrt{3}}$ |
| Exponential | $\frac{1}{\lambda}$ | $\frac{1}{\lambda}$ |
| Standard Normal | 0 | 1 |
| General Normal | $\mu$ | $\sigma$ |
| Chi-Square | $n$ | $\sqrt{2n}$ |

## The Standard Normal Distribution $\Phi(z)$

| $z$ | $\Phi(z)$ | $z$ | $\Phi(z)$ |
|---|---|---|---|
| 0.0 | .5000 | -1.2 | .1151 |
| -0.1 | .4602 | -1.4 | .0808 |
| -0.2 | .4207 | -1.6 | .0548 |
| -0.3 | .3821 | -1.8 | .0359 |
| -0.4 | .3446 | **-2.0** | **.0228** |
| -0.5 | .3085 | -2.2 | .0139 |
| -0.6 | .2743 | -2.4 | .0082 |
| -0.7 | .2420 | -2.6 | .0047 |
| -0.8 | .2119 | -2.8 | .0026 |
| -0.9 | .1841 | -3.0 | .0013 |
| -1.0 | .1587 | -3.2 | .0007 |

( For example, $P(Z \leq -2.0) = \Phi(\mathbf{-2.0}) = \mathbf{2.28}\%$ )

## The $\chi^2_n$ - Table

| $n$ | $\alpha = 0.975$ | $\alpha = \mathbf{0.95}$ | $\alpha = 0.05$ | $\alpha = 0.025$ |
|---|---|---|---|---|
| 5 | 0.83 | 1.15 | 11.07 | 12.83 |
| 6 | 1.24 | 1.64 | 12.59 | 14.45 |
| 7 | 1.69 | 2.17 | 14.07 | 16.01 |
| 8 | 2.18 | 2.73 | 15.51 | 17.54 |
| 9 | 2.70 | 3.33 | 16.92 | 19.02 |
| **10** | 3.25 | **3.94** | 18.31 | 20.48 |
| 11 | 3.82 | 4.58 | 19.68 | 21.92 |
| 12 | 4.40 | 5.23 | 21.03 | 23.34 |
| 13 | 5.01 | 5.89 | 22.36 | 24.74 |
| 14 | 5.63 | 6.57 | 23.69 | 26.12 |
| 15 | 6.26 | 7.26 | 25.00 | 27.49 |

This Table shows $z_{\alpha,n}$ values such that $P(\chi^2_n \geq z_{\alpha,n}) = \alpha$ .

( For example, $P(\chi^2_{\mathbf{10}} \geq \mathbf{3.94}) = \mathbf{95}\%$ )

## The $T$ - distribution Table

| $n$ | $\alpha = 0.1$ | $\alpha = 0.05$ | $\alpha = 0.01$ | $\alpha = 0.005$ |
|---|---|---|---|---|
| 5 | 1.476 | 2.015 | 3.365 | 4.032 |
| 6 | 1.440 | 1.943 | 3.143 | 3.707 |
| 7 | 1.415 | 1.895 | 2.998 | 3.499 |
| 8 | 1.397 | 1.860 | 2.896 | 3.355 |
| 9 | 3.383 | 1.833 | 2.821 | 3.250 |
| **10** | 1.372 | 1.812 | 2.764 | **3.169** |
| 11 | 1.363 | 1.796 | 2.718 | 3.106 |
| 12 | 1.356 | 1.782 | 2.681 | 3.055 |
| 13 | 1.350 | 1.771 | 2.650 | 3.012 |
| 14 | 1.345 | 1.761 | 2.624 | 2.977 |
| 15 | 1.341 | 1.753 | 2.602 | 2.947 |

This Table shows $z_{\alpha,n}$ values such that $P(T_n \leq z_{\alpha,n}) = 1 - \alpha$ .

( For example, $P(T_{\mathbf{10}} \leq \mathbf{3.169}) = \mathbf{99.5}\%$ )